Consommation et Affaires commerciales Canada
Consumer and Corporate Affairs Canada

Bureau des brevets
Patent Office

Ottawa, Canada
K1A 0C9

(21) (A1) 2,078,716

(22) 1992/09/21

(43) 1994/03/22

5,072,3/75

(51) INTL.CL.[5] C12N-015/31; C12P-021/08; C12Q-001/68; C12N-001/21; C07K-013/00; C07K-015/12; C07H-021/00; A61K-039/40; A61K-039/108; G01N-033/569

(19)(CA) **APPLICATION FOR CANADIAN PATENT** (12)

(54) Attaching and Effacing Protein of Enterohemorrhagic E. Coli

(72) de Azavedo, Joyce - Canada ;
Brunton, James - Canada ;
Louie, Marie - Canada ;

(73) Mount Sinai Hospital Corporation - Canada ;

(57) 32 Claims

ABSTRACT OF THE DISCLOSURE

A DNA segment having a sequence encoding a protein associated with attaching and effacing activity of enterohemorrhagic E.coli, or an oligonucleotide fragment thereof is provided. The DNA segment permits selection of DNA and amino acid sequences unique to a protein associated with attaching and effacing activity of enterohemorrhagic E.coli. Novel DNA segments or proteins can thus be constructed which contain the unique DNA and amino acid sequences. The DNA segments or oligonucleotide fragments thereof may be use to detect the presence of a DNA segment having a sequence encoding a protein associated with attaching and effacing activity of enterohemorrhagic E.coli. Selected DNA segments or oligonucleotide fragments thereof may be used in methods for detecting pathogenic verotoxin-producing E.coli.

Title: Attaching and Effacing Protein of Enterohemorrhagic E.coli

FIELD OF THE INVENTION

The present invention relates to a DNA segment having a sequence encoding a protein associated with attaching and effacing activity of enterohemorrhagic E.coli, or an oligonucleotide fragment thereof. The DNA segment of the invention can be used to produce a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or a part thereof, by culturing a transformant host cell which includes a recombinant molecule comprising a DNA segment of the invention or an oligonucleotide fragment thereof and an expression control sequence operatively linked to the DNA segment.

The DNA segment of the invention also permits selection of DNA and amino acid sequences unique to a protein associated with attaching and effacing activity of enterohemorrhagic E.coli. Novel DNA segments or proteins can thus be constructed which contain the unique DNA and amino acid sequences. The invention also relates to uses of the DNA segments encoding a protein associated with attaching and effacing activity of enterohemorrhagic E.coli. and fragments thereof, and uses of a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or parts thereof.

BACKGROUND OF THE INVENTION

Enteropathogenic *Escherichia coli* (EPEC) colonize the intestine of humans (Ulshen, M., and Rollo J. (1980) N. Engl. J. Med. 302, 99-102) and experimental animals, (Tzipori, S., Robbins-Brown, R.M., Gonis, G., Hayes, J., Withers, M., and McCartney, E. (1985) Gut 26, 570-578) producing a characteristic lesion which has been called attaching and effacing (AE) (Moon, H.W., Whipp, S.C., Argenzio, R.A., Levin, M.M., and Gianella, R.A. 1983.

Infect. Immun. 41, 1340-1351). The lesion consists of an extremely close (10nm) approximation of the bacterial outer membrane to the intestinal epithelial cell membrane. The microvilli of the enterocyte are effaced and there is polymerization of actin in the cytoplasm subjacent to the area of bacterial attachment (Knutton, S., Baldwin, T., Williams, P.H., and McNeish, A.S. (1989) Infect. Immun. 57, 1290-1298). The AE lesion is also produced by enterohemorrhagic E.coli (EHEC) of serogroups 026, 0111 and 0157 (Tzipori, S., Wachsmuth, I., Smithers, J., and Jackson, C., (1988) Gastroenterology 94, 590-597; Tzipori, S., Wachsmuth, I.K., Chapman, C., Birner, R., Brittingham, J., Jackson, C., and Hogg, J. (1986) J. Infect. Dis. 154, 712-716).

Escherichia coli serotype O157:H7 is the most important member of the group of E.coli which produce verotoxins and is a major public health concern. Among the verotoxin-producing E.coli, serotype O157:H7 is most commonly associated with human disease. This serotype is a recognized cause of hemorrhagic colitis, occurring both sporadically and in major outbreaks, and is also a cause of the hemolytic uremic syndrome especially among children and the elderly. These organisms are probably acquired by humans from an animal source via contaminated meat and occasionally unpasteurized dairy products. Outbreaks of disease caused by O157:H7 have been associated with consumption of undercooked hamburgers and raw milk (Riley LW, Remis RS, Helgerson SD, et al. N Engl J Med 1983; 308:681-5; Borczyk AA, Karmali MA, Lior H, et al. (Letter). Lancet 1987; 1:98; Duncan L, Mai V, Carter A, et al. Outbreak of gastrointestinal disease - Ontario. Can Dis Wkly Rep 1987; 13-2:5-8).

A method that would identify these organisms in animal and food samples would be very useful for screening infected herds and contaminated food. Presently, the detection of verotoxin and DNA probes for verotoxin do not distinguish

verotoxin-producing E.coli which are important causes of human disease such as O157:H7 from those which, although present in animals, rarely if ever cause human disease.

Jerse et al. ((1990) Proc. Natl. Acad. Sci. USA 87, 7839-7843) and Donnenberg et al. ((1990) Infect. Immun. 58, 1565-1571) have identified a gene in the EPEC strain E2348/69 which was necessary but not sufficient for the formation of the AE lesion. This gene of approximately 3kb, which was designated *eae*, encodes a 94 kilodalton outer membrane protein (Jerse, A.E. and Kaper, J.B. (1991) Abst. Ann. Am. Soc. Microbiol. B-112, p.44). It was shown that E2348/69 mutants carrying transposon Tn*PhoA* insertions in the *eae* gene were unable to produce the AE lesion when incubated with HEp-2 cells (Donnenberg M., Calderwood, S., Donohue-Rolfe, A., Keusch, G.T., and Kaper, J. (1990) Infect. Immun. 58, 1565-1571). A 1kb *Stu*I-*Sal*I fragment which encompasses the central one third of the EPEC *eae* gene was found to hybridize with DNA isolated from bacteria of classical EPEC serogroups as well as with DNA isolated from EHEC of serogroups 026 and 0157 (Jerse, A., Yu, J., Tall, B., and Kaper, J. (1990) Proc. Natl. Acad. Sci. USA 87, 7839-7843). Although both EPEC and EHEC strains cause ultrastructurally similar lesions, as shown in gnotobiotic piglets (Moon, H.W., Whipp, S.C., Argenzio, R.A., Levin, M.M., and Gianella, R.A. 1983. Infect. Immun. 41, 1340-1351; Tzipori, S., Wachsmuth, I.K., Chapman, C., Birner, R., Brittingham, J., Jackson, C., and Hogg, J. (1986) J. Infect. Dis. 154, 712-716) and in tissue culture cells (Knutton, S., Baldwin, T., Williams, P.H., and McNeish, A.S. (1989) Infect. Immun. 57, 1290-1298), the two types of infection can be differentiated by anatomic site of involvement, suggesting differences in adherence factors.

SUMMARY OF THE INVENTION

The present invention provides a purified and isolated DNA segment having a sequence which codes for a protein

associated with attaching and effacing activity of enterohemorrhagic E.coli. In a preferred embodiment, a purified and isolated DNA segment is provided having a sequence which codes for a protein associated with attaching and effacing activity of enterohemorrhagic E.coli having an amino acid sequence which has substantial homology with the amino acid sequence as shown in the Sequence Listing as SEQ ID NO:1. Most preferably, the purified and isolated DNA segment has a sequence having substantial sequence homology with the nucleotide sequence as shown in the Sequence Listing as SEQ ID NO:1.

The invention also relates to a recombinant molecule adapted for transformation of a host cell comprising the DNA segment of the invention operatively linked to an expression control sequence. A transformant host cell including a recombinant molecule of the invention is also provided. Still further, this invention provides plasmids which comprise the recombinant molecules of the invention.

The present invention further relates to an avirulent strain of an enterohemorrhagic E.coli comprising an avirulent bacterial carrier strain transformed with a recombinant molecule of the invention, and a vaccine composition comprising a bacterial carrier strain transformed with a recombinant molecule of the invention.

The invention also provides a method of preparing a protein associated with attaching and effacing activity of enterohemorrhagic E.coli, utilizing a DNA segment of the invention. The method comprises culturing a transformant host cell including a recombinant molecule comprising a DNA segment of the invention and an expression control sequence operatively linked to the DNA segment, in a suitable medium until the protein is formed and thereafter isolating the protein.

The invention still further provides a substantially pure protein associated with attaching and effacing activity of enterohemorrhagic E.coli. In a preferred embodiment the protein has an amino acid sequence which has substantial homology with the amino acid sequence as shown in the Sequence Listing as SEQ ID NO:1.

The invention also relates to antibodies specific for an epitope of a protein of the invention, preferably a monoclonal antibody, and a method for preparing the antibodies.

A diagnostic kit for detecting enterohemorrhagic E.coli, preferably E.coli 0157 in a sample comprising a monoclonal antibody of the invention and directions for its use is also provided.

The invention also contemplates an immunoassay for the detection of enterohemorrhagic E.coli, preferably E.coli 0157, comprising contacting a sample suspected of containing enterohemorrhagic E.coli with a monoclonal antibody of the invention to form an immune complex; and determining the presence of the immune complex.

The DNA segments of the invention or oligonucleotide fragments of the DNA segments, allow those skilled in the art to construct nucleotide probes for use in the detection of nucleotide sequences in samples such as biological or food samples. The nucleotide probes may be used to detect nucleotide sequences that encode proteins related to or analogous to the attaching and effacing protein of the invention.

Accordingly, the invention provides a method for detecting the presence of a DNA segment having a sequence encoding a protein related to or analogous to a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or an oligonucleotide fragment thereof in a sample,

comprising contacting the sample with a nucleotide probe capable of hybridizing with the DNA segment or an oligonucleotide fragment thereof, to form a hybridization product, under conditions which permit the formation of the hybridization product, and assaying for the hybridization product.

The invention further provides a kit for detecting the presence of a DNA segment having a sequence encoding a protein related to or analogous to a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or an oligonucleotide fragment thereof in a sample, comprising a nucleotide probe capable of hybridizing with the DNA segment or an oligonucleotide fragment thereof, reagents required for hybridization of the nucleotide probe with the DNA segment or an oligonucleotide fragment thereof, and directions for its use.

The DNA segment of the invention also permits the identification and isolation, or synthesis, of nucleotide sequences which may be used as primers to amplify a DNA segment of the invention or an oligonucleotide fragment thereof, for example in the polymerase chain reaction (PCR).

Accordingly, the invention relates to a method of determining the presence of a DNA segment having a sequence encoding a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or a predetermined oligonucleotide fragment thereof in a sample, comprising treating the sample with primers which are capable of amplifying the DNA segment or the predetermined oligonucleotide fragment thereof, in a polymerase chain reaction to form amplified sequences, under conditions which permit the formation of amplified sequences, and, assaying for amplified sequences.

The invention further relates to a kit for determining the presence of a DNA segment having a sequence encoding a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or a predetermined oligonucleotide fragment thereof in a sample, comprising primers which are capable of amplifying the DNA segment or the predetermined oligonucleotide fragment thereof in a polymerase chain reaction to form amplified sequences, means for assaying the amplified sequences, and directions for its use.

The present inventors have designed specific nucleotide sequences for detecting pathogenic verotoxin-producing E.coli strains. The present invention therefore further provides a purified and isolated nucleotide sequence comprising the sequence shown in the Sequence Listing as SEQ ID NO:2; the sequence shown in the Sequence Listing as SEQ ID NO:3; the sequence shown in the Sequence Listing as SEQ ID NO:4, the sequence shown in the Sequence Listing as SEQ ID NO:5; the sequence shown in the Sequence Listing as SEQ ID NO:6; the sequence shown in the Sequence Listing as SEQ ID NO:7; the sequence shown in the Sequence Listing as SEQ ID NO:8; the sequence shown in the Sequence Listing as SEQ ID NO:9; a sequence having substantial homology thereto, or a fragment of the nucleotide sequence.

Accordingly, the present invention relates to a method for detecting verotoxin-producing E.coli of the serogroups 05, 026, 0103, 0111, 0118, 0145, and 0157 in a sample comprising contacting the sample with a SalI-StuI fragment comprising the sequence as shown in the Sequence Listing as SEQ ID NO:2 and the sequence as shown in the Sequence Listing as SEQ ID NO:3, under conditions which permit the fragment to hybridize with a complementary sequence in the sample to form a hybridization product, and assaying for the hybridization product.

The invention also contemplates a method for detecting verotoxin-producing E.coli serogroup 0157 in a sample comprising contacting the sample with a 0.5 kb fragment comprising the sequence as shown in the Sequence Listing as SEQ ID NO:4 and the sequence as shown in the Sequence Listing as SEQ ID NO:5, under conditions which permit the fragment to hybridize with a complementary sequence in the sample to form a hybridization product, and assaying for the hybridization product.

A method for detecting pathogenic verotoxin-producing E.coli strains of the serogroup 0157 in a sample is also provided comprising treating the sample with the nucleotide sequence P1EH shown in the Sequence Listing as SEQ ID NO:6 and the nucleotide sequence P2EH5 shown in the Sequence Listing as SEQ ID NO:7 in the polymerase chain reaction, under conditions which permit the formation of amplified sequences which terminate with the nucleic acid sequence of one of the primers and the complementary nucleic acid sequence of the other of the primers, and, assaying for the amplified sequences.

The invention further provides a kit for detecting verotoxin-producing E.coli of the serogroups 05, 026, 0103, 0111, 0118, 0145, and 0157 in a sample comprising a SalI-StuI fragment comprising the sequence as shown in the Sequence Listing as SEQ ID NO:2 and the sequence as shown in the Sequence Listing as SEQ ID NO:3, reagents required for hybridization of the fragment with a DNA segment encoding the attaching and effacing protein of enterohemorrhagic E.coli or an oligonucleotide fragment thereof, and directions for its use.

The invention also relates to a kit for detecting verotoxin-producing E.coli of the serogroup 0157 in a sample is comprising a 0.5 kb fragment (A3-B2) comprising the sequence as shown in the Sequence Listing as SEQ ID NO:4 and the sequence as shown in the Sequence Listing as

SEQ ID NO:5, reagents required for hybridization of the fragment with a DNA segment encoding the attaching and effacing protein of enterohemorrhagic E.coli or an oligonucleotide fragment thereof, and directions for its use.

The invention further relates to a kit for detecting pathogenic verotoxin-producing E.coli strains of the serogroup 0157 in a sample comprising the primers P1EH as shown in the Sequence Listing as SEQ ID NO:6 and P2EH as shown in the Sequence Listing as SEQ ID NO:7, respectively, reagents required for the amplification in a PCR reaction of sequences which terminate with the nucleic acid sequence of one of the primers and the complementary nucleic acid sequence of the other of the primers, and directions for its use.

DESCRIPTION OF THE DRAWINGS

The invention will now be described in relation to the drawings in which:

Figure 1 shows the DNA sequence of the EHEC *eae* gene and the deduced protein sequence starting at the start codon at 206 bp;

Figure 2 shows the homology between the deduced protein sequences of the EPEC and EHEC *eae* genes;

Figure 3 shows the amino acid sequence alignment of the predicted EPEC and EHEC proteins;

Figure 4 shows the results of colony blot hybridizations of various O serogroups using A) a central (C1-C2) 1 kb fragment and B) a 3' (A3-B2) 0.5 fragment of the O157:H7 eae gene;

Figure 5 shows the results of colony blot hybridizations of various O serogroups using a 0.45 kb fragment (P1EH-P2EH) of the O157:H7 eae gene;

Figure 6 shows the primer locations on the 3' region of EHEC eae gene;

Figure 7 shows the amplification products of VTEC and EPEC using specific primer sets A3-B2 (0.5 kb) and P1EH-P2EH (0.45 kb);

Figure 8 shows the results of colony blot hybridizations of serogroups O157 (VT+), O157 (VT-), O5, O26, O111, O113, O121, and O127 with [α^{32}P] dATP end-labeled with P2EH oligonucleotide; and

Figure 9 shows the alignment of the deduced amino acid sequences at the 3' end of: O157: EHEC eae O157:H7 protein; O111: EHEC eae O111:H8 protein; EPEC: EPEC eae O127:H6 protein; and INV : Y.pseudotuberculosis invasin.

DETAILED DESCRIPTION OF THE INVENTION

The present inventors have identified a chromosomal eae gene associated with attaching and effacing activity in enterohemorrhagic E.coli (EHEC). The EHEC eae gene was identified from EHEC serotype O157:H7 strain CL-8. The entire predicted structural gene was sequenced including 205 bp upstream and 118 bp downstream. Figure 1 and the sequence shown in the Sequence Listing as SEQ ID NO:1, show the DNA sequence of the EHEC *eae* gene and the deduced protein sequence starting at the start codon at 206 bp. Comparison of nucleotide sequence similarity between the EHEC and EPEC eae genes shows 97% homology in the first 2200 bp and only 59% homology in the last 800 bp (Figure 2). The amino acid sequence alignment of the predicted EPEC and EHEC proteins is shown in Figure 3. The N-terminal amino acids are highly conserved for the first 685 amino acids while for the C-terminal amino acids, only

short stretches can be aligned. There is conservation of cysteine residues at positions 859 and 933 of the alignment.

Both the EHEC and EPEC sequences show similarity to the Yersinia pseudotuberculosis invasin gene (Isberg, R.R. et al., (1987) Cell 50, p.769) with greatest divergence at the C-terminus. The C-terminal end of the Yersinia inv gene is associated with receptor binding (Leong et al, EMBO J 1990; 6:1979-89) and it is possible that the same applies to the eae gene products.

The inventors also probed verotoxin-producing E.coli (VTEC) (251 strains) of multiple O-serogroups with a highly conserved 1kb StuI-SalI central fragment from the EHEC eae gene using colony hybridization. Only the following VTEC serogroups (125 strains) were found to be positive: O5, O26, O103, O111, O118, O145, and O157. Using the polymerase chain reaction (PCR) a less conserved 0.5kb probe was generated from the 3' end of the EHEC eae gene. Among the 251 VTEC strains tested using colony hybridization, only those strains belonging to the O157 serogroup were positive. Additionally, the 3'end of eae genes of VTEC strains appear to have unique nucleotide sequences as determined on selected O-serogroups. These O-serogroups initially hybridized with the highly conserved central regions of the eae genes.

Therefore, the present inventors have shown that the central 1 kb fragment of the EHEC eae gene is homologous in a variety of O serogroups of verotoxin-producing E.coli (VTEC) strains but that there is diversity at the C-terminal end. The heterogeneity at the 3' end of eae genes allows for specific detection of clinically relevant VTEC strains and may be associated with differences in receptor binding.

Accordingly, the present invention provides a purified and isolated DNA segment having a sequence which codes for a protein associated with attaching and effacing activity of enterohemorrhagic E.coli.

In a preferred embodiment, a purified and isolated DNA segment is provided having a sequence which codes for a protein associated with attaching and effacing activity of enterohemorrhagic E.coli having an amino acid sequence which has substantial homology with the amino acid sequence as shown in the Sequence Listing as SEQ ID NO:1. Most preferably, the purified and isolated DNA segment has a sequence having substantial sequence homology with the nucleotide sequence as shown in the Sequence Listing as SEQ ID NO:1.

DNA segments of the present invention encoding a protein associated with attaching and effacing activity of EHEC or related or analogous sequences may be isolated and sequenced, by selectively amplifying the region of the eae gene using the polymerase chain reaction method and genomic DNA. It is possible to design synthetic oligonucleotide primers from the sequence shown in the Sequence Listing as SEQ ID NO:1 for use in PCR and for screening genomic libraries. An amplified fragment can be cloned and characterized by DNA sequence analysis. DNA segments of the present invention encoding the protein associated with attaching and effacing activity of EHEC may also be constructed by chemical synthesis and enzymatic ligation reactions using procedures known in the art.

It will be appreciated that the invention includes nucleotide or amino acid sequences which have substantial sequence homology with the nucleotide and amino acid sequences shown in the Sequence Listing as SEQ ID NO:1. The term "sequences having substantial sequence homology" means those nucleotide and amino acid sequences which have

slight or inconsequential sequence variations from the sequences disclosed in the Sequence Listing as SEQ ID NO:1 i.e. the homologous sequences function in substantially the same manner to produce substantially the same polypeptides as the actual sequences. The variations may be attributable to local mutations or structural modifications.

It will also be appreciated that a double stranded nucleotide sequence comprising a DNA segment of the invention as shown in Figure 1 or an oligonucleotide fragment thereof, hydrogen bonded to a complementary nucleotide base sequence (see Sequence Listing SEQ ID NO:1), an RNA made by transcription of this doubled stranded nucleotide sequence, and an antisense strand of a DNA segment of the invention or an oligonucleotide fragment of the DNA segment, are contemplated within the scope of the invention.

A number of unique restriction sequences for restriction enzymes are incorporated in the DNA segment identified in the Sequence Listing as SEQ ID NO:1, and these provide access to nucleotide sequences which code for polypeptides unique to the protein associated with attaching and effacing activity of EHEC. DNA sequences unique to the protein associated with attaching and effacing activity of EHEC or isoforms thereof, can also be constructed by chemical synthesis and enzymatic ligation reactions carried out by procedures known in the art.

The DNA segments of the invention or oligonucleotide fragments of the DNA segments, allow those skilled in the art to construct nucleotide probes for use in the detection of nucleotide sequences in samples such as biological specimens or food samples. A nucleotide probe may be labelled with a detectable marker such as a radioactive label which provides for an adequate signal and has sufficient half-life such as ^{32}P, ^{3}H, ^{14}C or the

like. Other labels which may be used include antigens that are recognized by a specific labelled antibody, fluorescent compounds, enzymes, antibodies specific for a labelled antigen, and chemiluminescent compounds. An appropriate label may be selected having regard to the rate of hybridization and binding of the probe to the nucleotide to be detected and the amount of nucleotide available for hybridization.

The nucleotide probes may be used to detect DNA segments having sequences that encode proteins related to or analogous to the protein associated with attaching and effacing activity of EHEC.

Accordingly the present invention also relates to a method for detecting the presence of a DNA segment having a sequence encoding a protein related to or analogous to a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or an oligonucleotide fragment thereof in a sample, comprising contacting the sample with a nucleotide probe capable of hybridizing with the DNA segment or an oligonucleotide fragment thereof to form a hybridization product, under conditions which permit the formation of the hybridization product, and assaying for the hybridization product. The nucleotide probe may be labelled with a detectable marker as described herein and the hybridization product may be assayed by detecting the detectable marker or the detectable change produced by the detectable marker.

The invention also provides a kit for detecting the presence of a DNA segment having a sequence encoding a protein related to or analogous to a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or an oligonucleotide fragment thereof in a sample, comprising a nucleotide probe capable of hybridizing with the DNA segment or an oligonucleotide fragment thereof, reagents required for hybridization of the nucleotide

probe with the DNA segment or an oligonucleotide fragment thereof, and directions for its use.

The DNA segment of the invention permits the identification and isolation, or synthesis of nucleotide sequences which may be used as primers to amplify a DNA segment of the invention or an oligonucleotide sequence thereof, for example in the polymerase chain reaction (PCR). The length and bases of the primers for use in the PCR are selected so that they will hybridize to different strands of the desired sequence and at relative positions along the sequence such that an extension product synthesized from one primer when it is separated from its template can serve as a template for extension of the other primer into a nucleic acid of defined length.

The primers may be prepared using techniques known in the art such as for example phosphotriester and phosphodiester methods or automated techniques. Restriction endonuclease digests may also be used as primers.

It will be appreciated that the primers may contain non-complementary sequences provided that a sufficient amount of the primer contains a sequence which is complementary to the DNA segment of the invention or oligonucleotide sequence thereof, which is to be amplified. Restriction site linkers may also be incorporated into the primers allowing for digestion of the amplified products with the appropriate restriction enzymes facilitating cloning and sequencing of the amplified product.

It will also be appreciated that the invention includes nucleotide sequences which have substantial sequence homology with the above-mentioned nucleotide probes and primers. Further, it will be appreciated that a double stranded nucleotide sequence comprising a nucleotide probe or primer, hydrogen bonded to a complementary nucleotide base sequence, an RNA made by transcription of this

doubled stranded nucleotide sequence, and an antisense strand of a nucleotide probe or primer, are contemplated within the scope of the invention.

In an embodiment of the invention a method of determining the presence of a DNA segment having a sequence encoding a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or a predetermined oligonucleotide fragment thereof in a sample, is provided comprising treating the sample with primers which are capable of amplifying the DNA segment or the predetermined oligonucleotide fragment thereof in a polymerase chain reaction to form amplified sequences, under conditions which permit the formation of amplified sequences and, assaying for amplified sequences.

The polymerase chain reaction refers to a process for amplifying a target nucleic acid sequence as generally described in Innis et al, Academic Press, 1990 and U.S. Patent 4,800,159. Generally, amplification of the target nucleic acid sequence may be accomplished by means of a pair of primers which flank the nucleic acid sequence to be amplified. The primers hybridize to opposite strands of the target sequence and DNA synthesis proceeds across the region between the primers, thereby doubling the amount of that DNA segment. Repeated cycles of denaturation, priming and extension permit rapid exponential amplification of the target sequence.

The amplified products can be isolated and distinguished based on their respective sizes using techniques known in the art. For example, after amplification, the DNA sample can be separated on an agarose gel and visualized, after staining with ethidium bromide, under ultra violet (UV) light. DNA may be amplified to a desired level and a further extension reaction may be performed to incorporate nucleotide derivatives having detectable markers such as radioactive labelled or biotin labelled nucleoside

triphosphates. The detectable markers may be analyzed by restriction and electrophoretic separation or other techniques known in the art.

The conditions which may be employed in the methods of the invention using PCR are those which permit hybridization and amplification reactions to proceed in the presence of DNA in a sample and appropriate complementary hybridization primers. Conditions suitable for the polymerase chain reaction are generally known in the art. Preferably, the PCR utilizes Taq polymerase (GeneAmp Kit, Perkin Elmer Cetus) as the polymerization agent and each cycle consists of the following: denaturation at 94°C X 1 min; annealing at 55°C X 1 min; and extension at 72°C X 1 min.

The invention still further provides a kit for determining the presence of a DNA segment encoding a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or a predetermined oligonucleotide fragment thereof in a sample comprising primers which are capable of amplifying the DNA segment or the predetermined oligonucleotide fragment thereof, in a polymerase chain reaction to form amplified sequences, means for detecting the amplified sequences, and directions for its use.

The present invention also relates to a method of detecting verotoxin-producing E.coli using the DNA segment of the invention and oligonucleotide fragments thereof. As hereinbefore mentioned the present inventors have found that a SalI-StuI fragment (C1-C2) comprising the sequences as shown in the Sequence Listing as SEQ ID NO:2 and SEQ ID NO:3, may be used to detect verotoxin-producing E.coli of the serogroups 05, 026, 0103, 0111, 0118, 0145, and 0157. A 0.5 kb fragment (A3-B2) comprising the sequences as shown in the Sequence Listing as SEQ ID NO:4 and SEQ ID

NO:5 has also been found to be useful for detecting verotoxin-producing E.coli serogroup 0157.

The present inventors have also designed specific nucleotide probes for detecting pathogenic verotoxin-producing E.coli strains of the serogroup 0157. In particular, the synthetic nucleotide probes P1EH and P2EH shown in the Sequence Listing as SEQ ID NO:6 and SEQ ID NO:7, respectively, were found to specifically detect only verotoxin-producing E.coli serotypes O157:H7 and O157:NM. Non verotoxin-producing O157:NM strains and other O157:H serotypes were not detected by P1EH and P2EH, making these probes ideal for detecting pathogenic O157 EHEC.

The synthetic nucleotide probes P1EH and P2EH shown in the Sequence Listing as SEQ ID NO:6 and SEQ ID NO: 7, respectively, were also used in PCR and were found to specifically detect only verotoxin-producing E.coli O157 serogroups. Non verotoxin-producing O157 serogroups were not detected by P1EH and P2EH, making these probes ideal for detecting pathogenic O157 EHEC.

The nucleotide sequences P1O and P2O shown in the Sequence Listing as SEQ ID NO:8 and SEQ ID NO:9, respectively, based on the O111:H8 sequence were found to detect verotoxin-producing O111:H8 and O111:NM but not O111:H11 strains indicating that there is further diversity in the carboxyl-terminal within the serogroup.

Accordingly, the present invention provides a purified and isolated nucleotide sequence comprising the sequence shown in the Sequence Listing as SEQ ID NO:2; the sequence shown in the Sequence Listing as SEQ ID NO:3; the sequence shown in the Sequence Listing as SEQ ID NO:4, the sequence shown in the Sequence Listing as SEQ ID NO:5; the sequence shown in the Sequence Listing as SEQ ID NO:6, the sequence shown in the Sequence Listing as SEQ ID NO:7; the sequence shown in the Sequence Listing as SEQ ID NO:8; the sequence shown

in the Sequence Listing as SEQ ID NO:9; a sequence having substantial homology thereto or a fragment of the nucleotide sequence. The location of the sequences of the sequence shown in the Sequence Listing as SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6 and SEQ ID NO:7 on the O157:H7 eae gene are shown in Figure 6.

The invention relates to a method for detecting verotoxin-producing E.coli of the serogroups 05, 026, 0103, 0111, 0118, 0145, and 0157 in a sample, comprising contacting the sample with a SalI-StuI fragment (C1-C2) comprising the sequence as shown in the Sequence Listing as SEQ ID NO:2 and the sequence as shown in the Sequence Listing as SEQ ID NO:3, under conditions which permit the fragment to hybridize with a complementary sequence in the sample to form a hybridization product, and assaying for the hybridization product.

The invention relates to a method for detecting verotoxin-producing E.coli serogroup 0157 in a sample comprising contacting the sample with a 0.5 kb fragment (A3-B2) comprising the sequence as shown in the Sequence Listing as SEQ ID NO:4 and the sequence as shown in the Sequence Listing as SEQ ID NO:5, under conditions which permit the fragment to hybridize with a complementary sequence in the sample to form a hybridization product, and assaying for the hybridization product.

Hybridization conditions which may be used in the method of the invention are known in the art and are described for example in Sambrook J, Fritch EF, Maniatis T. In: Molecular Cloning, A Laboratory Manual,1989. (Nolan C, Ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. The hybridization product may be assayed using techniques known in the art. The oligonucleotide fragments may be labelled with a detectable marker as described herein and the hybridization product may be assayed by

detecting the detectable marker or the detectable change produced by the detectable marker.

The invention still further relates to a method for detecting pathogenic verotoxin-producing E.coli strains of the serogroup 0157 in a sample comprising treating the sample with the nucleotide sequence P1EH shown in sequence the Sequence Listing as SEQ ID NO:6 and the nucleotide sequence P2EH5 shown in the Sequence Listing as SEQ ID NO:7 in a PCR reaction, under conditions which permit the formation of amplified sequences which terminate with the nucleic acid sequence of one of the primers and the complementary nucleic acid sequence of the other of the primers, and, assaying for the amplified sequences.

The PCR reaction and suitable conditions permitting formation of amplified sequences are described above.

A kit for detecting verotoxin-producing E.coli of the serogroups 05, 026, 0103, 0111, 0118, 0145, and 0157 in a sample is provided comprising a SalI-StuI fragment (C1-C2) comprising the sequence as shown in the Sequence Listing as SEQ ID NO:2 and the sequence as shown in the Sequence Listing as SEQ ID NO:3, reagents required for hybridization of the fragment with a DNA segment encoding the attaching and effacing protein of EHEC or an oligonucleotide fragment thereof, and directions for its use.

A kit for detecting verotoxin-producing E.coli of the serogroup 0157 in a sample is also provided comprising a 0.5 kb fragment (A3-B2) comprising the sequence as shown in the Sequence Listing as SEQ ID NO:4 and the sequence as shown in the Sequence Listing as SEQ ID NO:5, reagents required for hybridization of the fragment with a DNA segment encoding the attaching and effacing protein of EHEC or an oligonucleotide fragment thereof, and directions for its use.

The invention also contemplates a kit for detecting pathogenic verotoxin-producing E.coli strains of the serogroup 0157 in a sample comprising the primers P1EH and P2EH as shown in the Sequence Listing as SEQ ID NO:6 and as SEQ ID NO:7, respectively, reagents required for the amplification in a PCR reaction of sequences which terminate with the nucleic acid sequence of one of the primers and the complementary nucleic acid sequence of the other of the primers, and directions for its use.

The DNA segment of the present invention having a sequence which codes for a protein associated with attaching and effacing activity of EHEC, or an oligonucleotide fragment of the DNA segment including the nucleotide sequences of probes and primers described herein may be incorporated in a known manner into a recombinant molecule which ensures good expression of the protein or part thereof. In general, a recombinant molecule of the invention contains the DNA segment or an oligonucleotide fragment thereof of the invention and an expression control sequence operatively linked to the DNA segment or oligonucleotide fragment. The DNA segment of the invention or an oligonucleotide fragment thereof, may be incorporated into a plasmid vector, for example, pTZ18R.

The protein associated with attaching and effacing activity of EHEC or isoforms or parts thereof, may be obtained by expression in a suitable host cell using techniques known in the art. Suitable host cells include prokaryotic or eukaryotic organisms or cell lines, for example, E.coli JM 101 and E.coli LE 392. The protein or parts thereof may be prepared by chemical synthesis using techniques well known in the chemistry of proteins such as solid phase synthesis (Merrifield, 1964, J. Am. Chem. Assoc. 85:2149-2154) or synthesis in homogenous solution (Houbenweyl, 1987, Methods of Organic Chemistry, ed. E. Wansch, Vol. 15 I and II, Thieme, Stuttgart).

The protein associated with attaching and effacing activity of EHEC or isoforms or parts thereof, of the invention may be expressed in an avirulent bacterial carrier strain such as Salmonella and Shigella strains. Accordingly, in a further aspect of the invention an avirulent strain of an EHEC is provided comprising an avirulent bacterial carrier strain transformed with a recombinant molecule of the invention. The avirulent strain may provide the basis for a vaccine composition which may be useful for effecting immunity against diseases caused by EHEC, in particular E.coli serotype 0157:H7. The invention therefore also provides a vaccine composition comprising a bacterial carrier strain transformed with a recombinant molecule of the invention. The vaccine composition may be useful in effecting immunity against diseases such as hemorrhagic colitis and hemolytic uremic syndrome.

The vaccine compositions can be prepared by per se known methods for the preparation of pharmaceutically acceptable vaccines which can be administered to patients. The vaccine composition may be in an oral or injectable form and may include pharmaceutically acceptable vehicles. On this basis, the vaccine compositions include, albeit not exclusively, solutions of the bacterial carrier strain transformed with a recombinant molecule of the invention in association with one or more pharmaceutically acceptable vehicles or diluents, and contained in buffered solutions with a suitable pH and iso-osmotic with the physiological fluids.

The protein associated with attaching and effacing activity of EHEC or parts thereof, may be used to prepare monoclonal or polyclonal antibodies. Antibodies may be prepared which bind a distinct epitope in an unconserved region of the protein. Conventional methods can be used to prepare the antibodies. As to the details relating to the preparation of monoclonal antibodies reference can be made

to Goding, J.W., Monoclonal Antibodies: Principles and Practice, 2nd Ed., Academic Press, London, 1986. In general monoclonal antibodies are prepared by obtaining hybridomas derived from immortalizing antibody producing cells obtained from a mammal having been immunized with a protein, and screening the hybridomas for production of antibody which binds the isolated protein.

The polyclonal or monoclonal antibodies may be used to detect a protein associated with attaching and effacing activity of EHEC in various biological materials or food samples, for example they may be used in an Elisa, radioimmunoassay or histochemical tests. Thus, the antibodies may be used as diagnostic reagents to quantify the amount of protein associated with attaching and effacing activity in a sample.

The monoclonal and polyclonal antibodies may be labelled with a detectable marker including various enzymes, fluorescent materials, luminescent materials and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, biotin, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; and examples of suitable radioactive material include I^{125}, I^{131} or tritium.

The invention will be more fully understood by reference to the following examples. However, the examples are merely intended to illustrate embodiments of the invention and are not to be construed to limit the scope of the invention.

EXAMPLES

Example 1

Cloning and sequencing of eae gene homologue from EHEC strain 0157

The following materials and methods were used in the investigations illustrated in the example:

Bacteriological media

E.coli strains were grown in L-broth and L-agar (Lech, K. and Brent, R. (1989) in Current Protocols in Molecular Biology (Ausubel, F.M. et al. Eds.) John Wiley & Sons, New York) supplemented when necessary with 100ug/ml carbenicillin (Sigma Chemical Co., St.Louis, MO). For production of single stranded DNA, strains were grown in 2xYT broth (Sambrook, J., Fritch, E.F., and Maniatis, T. (1989) in Molecular Cloning, A Laboratory Manual (Nolan, C., Ed.) Cold Spring Harbor Laboratory Press, New York). Bacteriophage lambda was grown in lambda broth (Lech, K. and Brent, R. (1989) in Current Protocols in Molecular Biology (Ausubel, F.M. et al. Eds.) John Wiley & Sons, New York).

Bacterial strains and vectors

E.coli strains E2348/69 (EPEC), serotype O127:H6 and CL8 (EHEC), serotype O157:H7 were obtained from Dr. M. Karmali, The Hospital for Sick Children, Toronto, Canada, and have been previously described (Levine, M.M., Nataro, J.P., Karch, H., Baldini, M.M., Kaper, J.B., Black, R.E., Clements, M.L. and O'Brien, A.D. (1985) J. Infect. Dis. 152, 550-559; Karmali, M.A., Petric, M., Lim, C., Fleming, P.C., Arbus, G.S. and Lior, H. (1985) J. Infect. Dis. 151, 775-882; Sherman, P., Soni, R. and Karmali, M.A. (1988) Infect. Immun. 56, 756-761). *E.coli* JM101 (Messing,J. (1979) Recomb. DNA Tech. Bull. 2,43-48) was used as the host for plasmid pTZ18R (Mead, D.A., Szczesna-Skorupa, E., and Kemper, B. (1986) Protein Engineering. 1,67-74) and E.coli LE392 (Borck, K., Beggs, J.D., Brammar, W.J., Hopkins, A.S., and Murray, N.E. (1976) Mol. Gen. Genet. 146, 199-207; Murray, N.E.,Brammar, W.J., and Murray, K. (1977) Mol. Gen. Genet.

150,53-61) was used for propagation of bacteriophage lambda GEM-11 (Promega, Madison, WI). Bacteriophage M13K07 (Rokeach, L., Haselby, J. and Ahoch, S. (1988) Proc. Nat. Acad. Sci. USA. 85, 4832-4836) was used as a helper to produce single stranded DNA from pTZ18R and recombinant derivatives.

DNA manipulation and sequencing

Plasmid DNA was prepared by the method of Birnboim and Doly (Birnboim, H.C. and Doly, J. (1979) Nucl. Acids Res. 7, 1513-1523). Bacterial genomic DNA was extracted using hexadecyltrimethyl ammonium bromide precipitation (Murray, M.G., and Thompson, W.F. (1980) Nucl. Acids Res. 8,4321-4325). Genomic DNA isolated from the EHEC strain CL8 was partially digested with *Sau*3A and fragments were size-fractionated by sucrose density gradient centrifugation (Sambrook, J., Fritch, E.F., and Maniatis, T. (1989) in Molecular Cloning, A Laboratory Manual (Nolan, C., Ed.) Cold Spring Harbor Laboratory Press, New York). The desired fragments (14-23kb) were isolated and ligated to dephosphorylated lambda Gem-11 *Bam*Hlarms (Promega, Wisconsin, MA] and packaged using the Packagene System extracts (Promega,Wisconsin, MA). Bacteriophage particles were propagated in E.coli LE392 and plated for plaque isolation.

The 1 Kb *Stu*I-*Sal*I fragment described by Jerse et al (Jerse, A., Yu, J., Tall, B., and Kaper, J. (1990) Proc. Natl. Acad. Sci. USA 87, 7839-7843) was used as a probe for the *eae* gene. It was produced by amplification of the central region of the *eae* gene of strain E2348/69 by the polymerase chain reaction (Sambrook, J., Fritch, E.F., and Maniatis, T. (1989) in Molecular Cloning, A Laboratory Manual (Nolan, C., Ed.) Cold Spring Harbor Laboratory Press, New York) using *Taq* polymerase (Perkin Cetus Elmer Corp., Norwalk, CT). Primers flanking the *Stu*I and *Sal*I sites were designed using the nucleotide sequence of the eae gene of E2348/69 deposited in Genbank by Jerse et al.

(Jerse, A., Yu, J., Tall, B., and Kaper, J. (1990) Proc. Natl. Acad. Sci. USA 87, 7839-7843). The upstream primer was ATGGAATTCTCGTCACAGTTGCAGGCCTGGT (2241-2263 of the eae sequence) and the downstream primer was ATGGAATTCCGAAGTCTTATCA- GCCGTAAAGT (3350-3328 of the *eae* sequence) with *Eco*R1 recognition sequences included at the 5'end of each primer. The fragment was labelled with α^{32}P dATP using a random priming kit (Boehringer Mannheim).This probe was used in plaque hybridizations to further identify clones containing the putative *eae* sequence.

Fragments which hybridized with the probe were subcloned in pTZ18R (Sambrook J, Fritch EF, Maniatis T. In: Molecular cloning, A Laboratory Manual,1989. (Nolan C, Ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.). Nested deletions were constructed by sequential digestion with exonuclease III and S1 nuclease using the Erase-a-base System (Promega, Madison, WI). Single stranded templates were prepared from fragments cloned in pTZ18R by superinfection with bacteriophage M13K07. DNA sequencing was performed using the Sanger dideoxy chain termination method (Sanger, F., Niklen, S. and Coulson, A.R.(1977) Proc.Nat.Acad. Sci. USA. 74, 5463-5468) with Sequenase Version 2.0 (United States Biochemicals, Cleveland, Ohio). The entire sequence of a 3.5Kb region including the *eae* gene was determined using oligonucleotide primers where necessary. Sequence analysis was performed using the Wisconsin Genetics Computer Group sequence analysis software package version 6.0 (Devereux, J., Haeberli, P., and Smithies, O. (1984) Nucleic Acids Res. 12, 387-395).

Nucleotide sequence and accession number

The DNA sequence data was submitted to the EMBL database and assigned the accession number X60439.

Of the plaques screened, several hybridized to the EPEC *eae* probe on primary screening and one designated λGEM-11

D4-4 with an insert of 14Kb, was chosen for further study. The fragments carrying the EHEC *eae* homologue were identified by hybridization to the 1Kb EPEC probe. The entire predicted structural gene was sequenced including 205 bp upstream and 118 bp downstream. Figure 1 shows the DNA sequence of the EHEC eae gene and the deduced protein sequence starting at the start codon at 206 bp. The end of the open reading frame is denoted by an asterisk. The sequence was compared with that of the EPEC *eae* gene. The EPEC and EHEC sequences are virtually identical for the first 2200 bp of the structural gene and for approximately 200 bp upstream of the start site. However, there is considerable divergence in the last 800 bp where the similarity is only 59% (Figure 2).

The amino acid sequence alignment of the predicted EPEC and EHEC proteins is shown in Figure 3. The positions of the terminal cysteine residues are denoted by asterisks. The N-terminal amino acids are highly conserved for the first 685 amino acids while for the C- terminal amino acids, only short stretches can be aligned. There is conservation of cysteine residues at positions 859 and 933 of the alignment.

Jerse et al. first noted the similarity between the central regions of eae gene of EPEC and the *inv* gene of *Yersinia pseudotuberculosis* (Jerse AE, Yu J, Tall BD, Kaper, J.B. Proc Natl Acad Sci USA 1990; 87:7839-43). The invasin protein, a product of the *inv* gene is an outer membrane protein of *Y. pseudotuberculosis* which has been extensively studied by Isberg et al. (Isberg, R.R., Voorhis, D.L. and Falkow, S. (1987) Cell. 50,769-778). When expressed in E.coli K12, the *inv* gene is necessary and sufficient for invasion of HEp-2 cells in tissue culture (Isberg, R.R., and Falkow, S. (1985) Nature. 317, 262-264). The invasin protein has been divided into three domains on the basis of mutational analysis and the study of fusion proteins (Isberg, R.R. (1989) Mol. Microbiol.

3,1449-1453). The amino terminus is important for export beyond the inner membrane while the central region is critical for stable localization in the outer membrane. The last 192 residues at the carboxyl terminus have been shown to be responsible for binding to β1 integrin receptors on the eukaryotic cell surface (Leong, J.M., Fournier, S., and Isberg, R.R. (1990) EMBO. Jnl. 9,1979-1989). This domain also carries the epitopes recognized by neutralizing monoclonal antibodies. Examination of the two sequences published for the *inv* genes of *Y. enterocolitica* (Miller, V.L. and Falkow, S. (1988) Infect. Immun. 56,1242-1248) and the *Y. pseudotuberculosis* (Isberg RR, Voorhis DL, and Falkow S. Cell 1987; 50:769-78) showed that there was similarity to the *eae* genes in the central one third of the sequence. The predicted protein products of the two *inv* genes also possess a subterminal cysteine residue with a second approximately 80 amino acids upstream. This suggests that the carboxyl termini of all four proteins contain a disulfide loop which may be necessary for biologic activity.

On the basis of the similarities between the inv and *eae* genes, the carboxyl termini of the *eae* proteins are predicted to be the receptor binding domains. It is likely that the dissimilarities in this region of the molecules are responsible for antigenic variation and possibly for differences in receptor binding specificity. Sherman et al (Sherman, P., Cockerill III, F., Soni, R., and Brunton, J. (1991) Infect. Immun. 59,890-899) showed that antiserum raised against a 94 kilodalton outer membrane protein of an O157:H7 strain inhibited adherence and AE lesion formation by the homologous EHEC strain but had no effect on the adherence of the EPEC strain E2348/69. Differences in receptor specificity could explain the differences observed in intestinal colonization by EPEC and EHEC strains. EPEC colonize the entire intestine in animal models and the upper small bowel of humans. In contrast,EHEC strains colonize the cecum and large

intestine in animal models while in humans they cause intense proximal colonic inflammation, characteristic of hemorrhagic colitis (Knutton, S., Baldwin, T., Williams, P.H., and McNeish, A.S.(1989) Infect. Immun. 57, 1290-1298; Tzipori, S., Wachsmuth, I., Smithers, J., and Jackson, C., (1988) Gastroenterology 94, 590-597; Tzipori, S., Wachsmuth, I.K., Chapman, C., Birner, R., Brittingham, J., Jackson, C., and Hogg, J. (1986) J. Infect.Dis. 154, 712-716; Riley, L. (1987) Annu. Rev. Microbiol. 41, 383-407; Karmali, M. (1989) Clin. Microbiol. Rev. 2, 15-38). Recent reports suggest that adhesions encoded by the high molecular weight plasmids of EPEC and EHEC strains are important in determining the specificity of adherence to tissue culture cells (Baldini, M.M., Kaper, J.B., Levine, M.M., Candy, D.C.A., and Moon, H.W. (1983) J. Ped. Gastroenterol. Nutr. 2,534-538; Karch, H., Hessemann, J., Laufs, R. , O'Brien, A.D., Tacket, C.O., and Levine, M.M. (1987) Infect. Immun. 55,455-461; Cantey, J.R., and Moseley, S.L. (1991) Infect. Immun. 59, 3924-3929; Jerse, A.E., Gicquelais, K.G., and Kaper, J.B. (1991) Infect.Immun. 59, 3869-3875). It remains to be seen whether differences in the C-terminal regions of the eae gene products of different EPEC and EHEC strains might result in alterations of receptor specificity.

EXAMPLE 2

Serotype Distribution of the EHEC eae Gene Among Verotoxin-Producing E.coli.

Probes based on the EHEC eae gene were used to study the serotype distribution of the eae homologue in a collection of verotoxin-producing E.coli isolated from human animal and food sources using colony hybridization. Colony hybridizations were carried out using standard techniques under highly stringent conditions (Sambrook J, Fritch EF, Maniatis T. In: Molecular cloning, A laboratory Manual,1989. (Nolan C, Ed.), Cold Spring Harbor Laboratory

Press, Cold Spring Harbor, NY.). DNA for probes was made by PCR amplification using strain O157:H7 or strain O111:H8. The primers used to amplify the probes are set forth in Table 1 and are also shown in the Sequence Listing as SEQ ID NOS: 2 to 9. DNA fragments used as probes were labeled with [α^{32}P] dATP using the Random Primed DNA Labeling Kit (Boehringer Mannheim). Oligonucleotide probes were end-labeled with [γ^{32}P] dATP using T4 polynucleotide kinase (Sambrook J, Fritch EF, Maniatis T. In: Molecular cloning, A laboratory Manual,1989. (Nolan C, Ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.).

DNA for sequencing was made by PCR amplification using primers based on sequences within the O157:H7 eae gene. Sequencing was carried out using the ds DNA Cycle Sequencing System (BRL Life Technologies) according to the manufacturer's instructions.

The results of the colony blot hybridizations using probes based on the O157:H7 EHEC eae gene are shown in Table 2 and Figure 4. The results show that whereas the central 1 kb fragment C1-C2 hybridized to a variety of the O serogroups of VTEC tested, a 0.5 kb fragment A3-B2 in the carboxyl-terminal region only hybridized to O157 strains. Eae homologues among VTEC were detected in serogroups O157, O111, O26, O5, O103, O145 and O118 but not O113 and O156. A 0.5 kb fragment (A3-B2) derived from the less conserved 3' region hybridized with serogroups O157 and occasional O111 strains. However, a more specific 0.45 kb probe (P1EH-P2EH) only hybridized with VTEC of serogroup O157 (Figure 5).

Example 3

Oligonucleotide primers specific to the EHEC eae gene were used in polymerase chain reaction assays to test 150 VTEC of diverse serogroups which were isolated from food, animal and human specimens. In particular, previously

characterized and serotyped isolates of verotoxin-producing E.coli and enteropathogenic E.coli strains derived from human, food and animal sources were contributed by A.Borczyk, C. Gyles and M.Karmali, The Hospital for Sick Children, Toronto, Canada. Several colonies were selected, resuspended in 50µl PCR buffer (X1) (Perkin Elmer Cetus) and boiled for 10 mins prior to use as template DNA in PCR amplifications. The PCR reaction conditions were as follows: 50µl reactions were set up using 1mM deoxynucleotide triphosphates, 1X PCR Buffer, 1 unit Taq polymerase (GeneAmp Kit, Perkin Elmer Cetus), 20 pmol of each primer and 1µl sample preparation. Each cycle (X40) consisted of the following:

denaturation @ 94°C X 1 min
annealing @ 55°C X 1 min
extension @ 72°C X 1 min.

Primers based on the O157:H7 sequence and the O111 sequence, (primers P1O and P2O) were used to amplify DNA from the various serogroups. The primers are shown in Table 1. The location of the A3, B2, P1EH, and P2EH primers on the O157:h7 eae gene are shown in Figure 6.

Colony blot hybridization using [γ^{32}P] dATP oligonucleotide probes was used to confirm the specificity of the primers. Colony blot hybridizations were performed as described in Example 2.

Verotoxin-producing E.coli of serogroups O157 and O111 strains were detected by PCR with primers A3-B2, giving a 0.5 kb fragment. Both primer sets A3-B2 and P1EH-P2EH did not detect non verotoxin-producing E.coli of the serogroups O157 and O111. Primer set P1EH-P2EH specifically detected only O157 verotoxin positive strains giving a 0.45 kb fragment (Table 3, Figure 7). Figure 7 shows the amplification products of VTEC and EPEC using specific primer sets A3-B2 (0.5 kb) and PIEH-P2EH (0.45kb) and in particular the following:

LANES 1,20:MW STANDARDS
LANE 3:O157:H7+
LANE 5:O111:H8+
LANE 7:O111:NM+
A3-B2 PRIMERS
LANE 8:O5:NM -
LANE 9:O127:H6-
LANE 13:O26:H11-
LANES 10,11:NEG CONTROLS
LANE 14:O157:H7+
LANE 15:O111:H8-
P1EH-P2EH PRIMERS
LANE 16:O111:NM-
LANE 19:NEG CONTROL

Colony blot hybridization using [γ^{32}P] dATP oligonucleotide probes confirmed the specificity of primers P1EH-P2EH data obtained from the PCR. Only verotoxin-producing O157 serogroups hybridized with P1EH and P2EH (Figure 8).

Attempts to amplify DNA from serogroups O5, O26 and O118 using primers based on the O157:H7 sequence were unsuccessful. The only serogroup which gave a PCR amplification product was an O111:H8 strain using primers 2214 and 3011. Comparisons of the C-terminal ends of the eae genes from an O157:H7, O111:H8, an EPEC strain and the Y. pseudotuberculosis inv gene are shown in Figure 9.

Based on the O111 sequence, primers P1O and P2O were designed which gave a PCR product for O111:NM and O111:H8 verotoxin-producing (VT+) strains but not O111:H11 indicating that there are differences among (VT+) O111 serotypes. However, non verotoxin-producing (VT-) O111:H2 and O111:H12 strains were negative.

The present invention has been described in detail and with particular reference to the preferred embodiments; however, it will be understood by one having ordinary

skill in the art that changes can be made thereto without departing from the spirit and scope thereof.

Forming part of the present disclosure are the appended sequence listings.

TABLE 1

PRIMERS

C1 5'ATG GAA TTC TCG TCA CAG TTG CAG GCC TGG T 3'sense

C2 5'ATG GAA TTC CGA AGT CTT ATC AGC CGT AAA GT 3'antisense

A3 5'ATG GAA TTC GGA TGT TCA ACG GTA AGT CT 3' sense

B2 5'ATG GAA TTC ACA ATA CCG TCT GTG TGG AT 3' antisense

2214 5'ATG GGA TCC CCG ATA TCG CAG TAG CAA ATG GTA AGG AT 3'sense

3011 5'ATG AAG CTT GAA TTC TCT ACA CAA ACC GCA T 3' antisense

P1EH 5'AAG CGA CTG AGG TCA CT 3' sense

P2EH 5'ACG CTG CTC ACT AGA TGT 3' antisense

P1O* 5'AGA CCT AGG TTA CAT TT 3' sense

P2O* 5'TAT TTT ATC AGC TTC AGT 3' antisense

TABLE 2

SEROGROUP DISTRIBUTION OF EAE-HOMOLOGUES AMONG VEROTOXIN-PRODUCING E.COLI

	PROBE		
	CENTRAL	3' REGION	
	C1-C2	A3-B2	P1EH-P2EH
SEROGROUP	(1.0kb)	(0.5kb)	(0.45 kb)
O157	67 / 67	18 / 18	7 / 7
O111	11 / 11	2 / 10*	0 / 5
O26	12 / 12	0 / 7	0 / 2
O5	7 / 7	0 / 6	0 / 1
O103	4 / 4	0 / 3	0 / 1
O145	5 / 5	0 / 13	-
O118	1 / 1	0 / 1	0 / 2
O113	0 / 5	0 / 4	0 / 1
O156	0 / 4	0 / 4	-
O127:H6 (E2348)	1 / 1	0 / 1	0 / 1
015:H- (RDEC1)	1 / 1	0 / 1	0 / 1

TABLE 3

DETECTION OF O157:H7 E.COLI BY PCR USING SPECIFIC PRIMERS

	PRIMER SET	
	A3-B2	P1EH-P2EH
SEROGROUPS	#Pos / #Tested	#Pos / #Tested
O157 (VT+) (H7,NM)	18 / 18	18 / 18
O157 (VT-)	0 / 8	0 / 8
O111 (VT+) (NM,H8,H11)	17 / 18	0 / 12
O111 (VT-) (H2,H12)	0 / 2	0 / 2
O26	0 / 6	0 / 2
VTEC*	0 / 27	0 / 16
ETEC**	0 / 5	0 / 3
EPEC***	0 / 4	0 / 1
EIEC (O136)	0 / 1	0 / 1

* VTEC (O1, O2, O5, O7, O16, O18, O22, O38, O40, O42, O91, O103, O113, O117, O118, O121, O145, OR, O7)

** ETEC (O6, O115, O148, O153)

***EPEC (O15, O119, O126, O127)

SEQUENCE LISTING

1. GENERAL INFORMATION
(a) APPLICANT: Joyce de Azavedo, James Brunton and Marie Louie
(b) TITLE OF INVENTION: Attaching and Effacing Protein of Enterohemorrhagic E.coli
(c) NUMBER OF SEQUENCES: 9

2. CORRESPONDENCE ADDRESS:
(a) ADDRESSEE: Linda M. Kurdydyk, Bereskin & Parr
(b) STREET: Box 401, 40 King Street West
(c) CITY: Toronto, Ontario
(d) COUNTRY: Canada
(e) ZIP CODE: M5H 3Y2

3. COMPUTER READABLE FORM:
(a) MEDIUM TYPE: Floppy disk
(b) COMPUTER: IBM PC compatible
(c) OPERATING SYSTEM: PC-DOS/MS-DOS
(d) SOFTWARE: DOS Text File

4. CURRENT APPLICATION DATA:
(a) APPLICATION NUMBER: Unknown
(b) FILING DATE: Unknown
(c) CLASSIFICATION: Unknown

5. ATTORNEY/AGENT INFORMATION:
(a) NAME: Linda M. Kurdydyk
(b) REGISTRATION NUMBER: 34,971
(c) REFERENCE NUMBER/DOCKET NUMBER: 3153-057/LMK

6. TELECOMMUNICATION INFORMATION:
(a) TELEPHONE: (416) 364-7311
(b) TELECOPY: (416) 361-1398

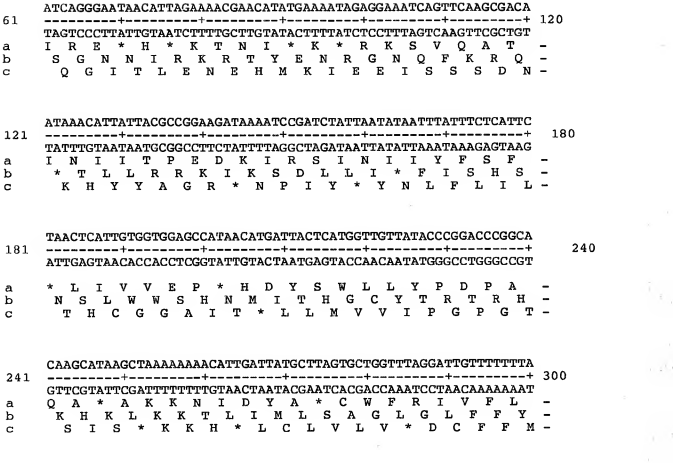
7. INFORMATION FOR SEQ ID NO:1:

(a) SEQUENCE CHARACTERISTICS:
(i) LENGTH: 3131 base pairs
(ii) TYPE:nucleic acid
(iii) STRANDEDNESS: double
(iv) TOPOLOGY: linear

(b) MOLECULAR TYPE: DNA (genomic)

(c) SEQUENCE DESCRIPTION: SEQ ID NO:1:

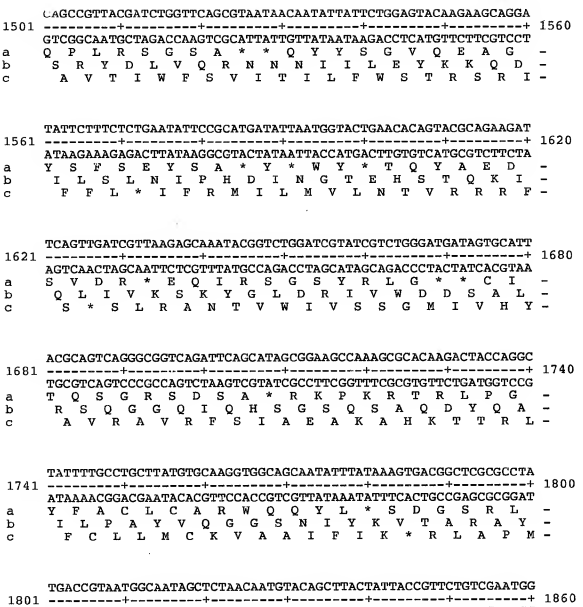
```
     TCGAGAATGAAATAGAAGTCGTTGTTAAGTCAATGGAAAACCTGTATTTGGTATTACATA
1    --------+---------+---------+---------+---------+---------+ 60
     AGCTCTTACTTTATCTTCAGCAACAATTCAGTTACCTTTTGGACATAAACCATAATGTAT
a:    S  R  M  K  *  K  S  L  L  S  Q  W  K  T  C  I  W  Y  Y  I  -
b:     R  E  *  N  R  S  R  C  *  V  N  G  K  P  V  F  G  I  T  *  -
c:      E  N  E  I  E  V  V  V  K  S  M  E  N  L  Y  L  V  L  H  N  -
```

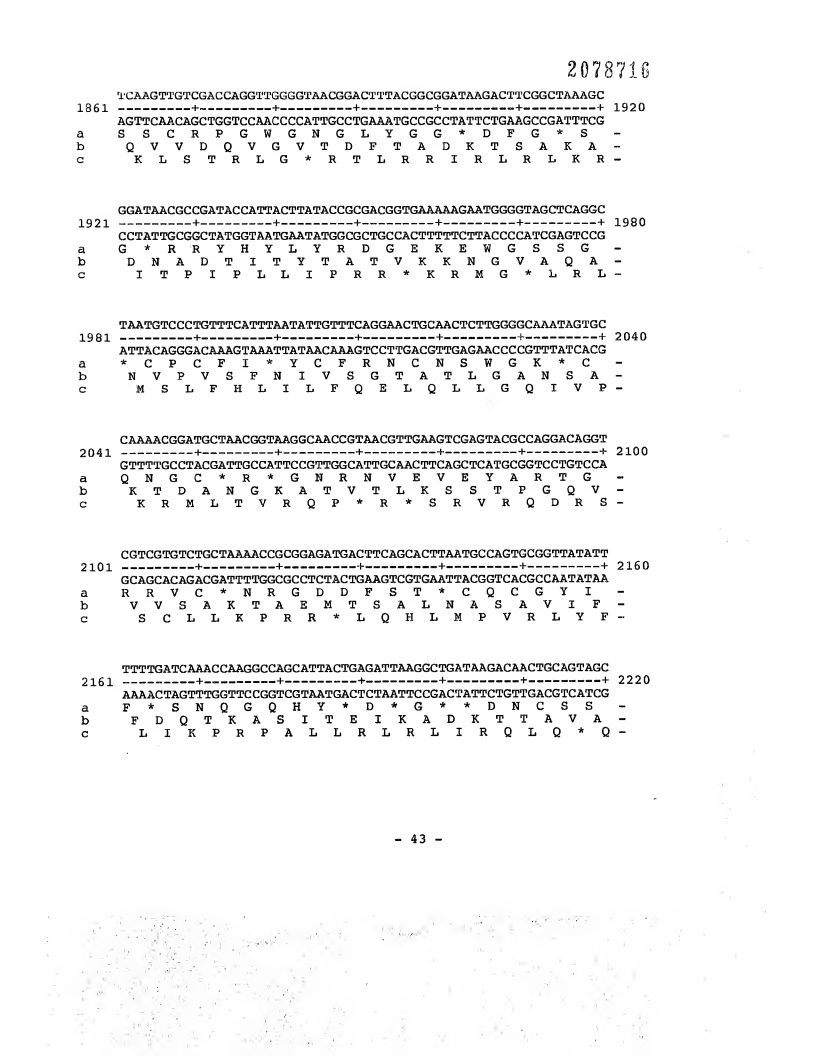

ATCAGGGAATAACATTAGAAAACGAACATATGAAAATAGAGGAAATCAGTTCAAGCGACA
TAGTCCCTTAATTGTAATCTTTTGCTTGTATACTTTTATCTCCTTTAGTCAAGTTCGCTGT
a I R E * H * K T N I * K * R K S V Q A T -
b S G N N I R K R T Y E N R G N Q F K R Q -
c Q G I T L E N E H M K I E E I S S S D N -
ATAAACATTATTACGCCGGAAGATAAAATCCGATCTATTAATATAATTTATTTCTCATTC
TATTTGTAATAATGCGGCCTTCTATTTTAGGCTAGATAATTATATTTAAATAAAGAGTAAG
a I N I I T P E D K I R S I N I I Y F S F -
b * T L L R R K I K S D L L I * F I S H S -
c K H Y Y A G R * N P I Y * Y N L F L I L -
TAACTCATTGTGGTGGAGCCATAACATGATTACTCATGGTTGTTATACCCGGACCCGGCA
ATTGAGTAACACCACCTCGGTATTGTACTAATGAGTACCAACAATATGGGCCTGGGCCGT
a * L I V V E P * H D Y S W L L Y P D P A -
b N S L W W S H N M I T H G C Y T R T R H -
c T H C G G A I T * L L M V V I P G P G T -
CAAGCATAAGCTAAAAAAAACATTGATTATGCTTAGTGCTGGTTTAGGATTGTTTTTTTA
GTTCGTATTCGATTTTTTTTGTAACTAATACGAATCACGACCAAATCCTAACAAAAAAAT
a Q A * A K K N I D Y A * C W F R I V F L -
b K H K L K K T L I M L S A G L G L F F Y -
c S I S * K K H * L C L V L V * D C F F M -
TGTTAATCAGAATTCATTTGCAAATGGTGAAAATTATTTTAAATTGGGTTCGGATTCAAA
ACAATTAGTCTTAAGTAAACGTTTACCACTTTTAATAAAATTTAACCCAAGCCTAAGTTT
a C * S E F I C K W * K L F * I G F G F K -
b V N Q N S F A N G E N Y F K L G S D S K -
c L I R I H L Q M V K I I L N W V R I Q N -
ACTGTTAACTCATGATAGCTATCAGAATCGCCTTTTTTATACGTTGAAAACTGGTGAAAC
TGACAATTGAGTACTATCGATAGTCTTAGCGGAAAAAATATGCAACTTTTGACCACTTTG
a T V N S * * L S E S P F L Y V E N W * N -
b L L T H D S Y Q N R L F Y T L K T G E T -
c C * L M I A I R I A F F I R * K L V K L -

```
     TGTTGCCGATCTTTCTAAATCGCAAGATATTAATTTATCGACGATTTGGTCGTTGAATAA
421  ---------+---------+---------+---------+---------+---------+ 480
     ACAACGGCTAGAAAGATTTAGCGTTCTATAATTAAATAGCTGCTAAACCAGCAACTTATT
a    C  C  R  S  F  *  I  A  R  Y  *  F  I  D  D  L  V  V  E  *   -
b     V  A  D  L  S  K  S  Q  D  I  N  L  S  T  I  W  S  L  N  K  -
c      L  P  I  F  L  N  R  K  I  L  I  Y  R  R  F  G  R  *  I  S -


     GCATTTATACAGTTCTGAAAGCGAAATGATGAAGGCCGCGCCTGGTCAGCAGATCATTTT
481  ---------+---------+---------+---------+---------+---------+ 540
     CGTAAATATGTCAAGACTTTCGCTTTACTACTTCCGGCGCGGACCAGTCGTCTAGTAAAA
a    A  F  I  Q  F  *  K  R  N  D  E  G  R  A  W  S  A  D  H  F   -
b     H  L  Y  S  S  E  S  E  M  M  K  A  A  P  G  Q  Q  I  I  L  -
c      I  Y  T  V  L  K  A  K  *  *  R  P  R  L  V  S  R  S  F  C -


     GCCACTCAAAAAACTTCCCTTTGAATACAGTGCACTACCACTTTTAGGTTCGGCACCTCT
541  ---------+---------+---------+---------+---------+---------+ 600
     CGGTGAGTTTTTTGAAGGGAAACTTATGTCACGTGATGGTGAAAATCCAAGCCGTGGAGA
a    A  T  Q  K  T  S  L  *  I  Q  C  T  T  T  F  R  F  G  T  S   -
b     P  L  K  K  L  P  F  E  Y  S  A  L  P  L  L  G  S  A  P  L  -
c      H  S  K  N  F  P  L  N  T  V  H  Y  H  F  *  V  R  H  L  L -


     TGTTGCTGCAGGTGGTGTTGCTGGTCACACGAATAAACTGACTAAAATGTCCCCGGACGT
601  ---------+---------+---------+---------+---------+---------+ 660
     ACAACGACGTCCACCACAACGACCAGTGTGCTTATTTGACTGATTTTACAGGGGCCTGCA
a    C  C  C  R  W  C  C  W  S  H  E  *  T  D  *  N  V  P  G  R   -
b     V  A  A  G  G  V  A  G  H  T  N  K  L  T  K  M  S  P  D  V  -
c      L  L  Q  V  V  L  L  V  T  R  I  N  *  L  K  C  P  R  T  * -


     GACCAAAAGCAACATGACCGATGACAAGGCATTAAATTATGCGGCACAACAGGCGGCGAG
661  ---------+---------+---------+---------+---------+---------+ 720
     CTGGTTTTCGTTGTACTGGCTACTGTTCCGTAATTTAATACGCCGTGTTGTCCGCCGCTC
a    D  Q  K  Q  H  D  R  *  Q  G  I  K  L  C  G  T  T  G  G  E   -
b     T  K  S  N  M  T  D  D  K  A  L  N  Y  A  A  Q  Q  A  A  S  -
c      P  K  A  T  *  P  M  T  R  H  *  I  M  R  H  N  R  R  R  V -


     TCTCGGTAGCCAGCTTCAGTCGCGATCTCTGAACGGCGATTACGCGAAAGATACCGCTCT
721  ---------+---------+---------+---------+---------+---------+ 780
     AGAGCCATCGGTCGAAGTCAGCGCTAGAGACTTGCCGCTAATGCGCTTTCTATGGCGAGA
a    S  R  *  P  A  S  V  A  I  S  E  R  R  L  R  E  R  Y  R  S   -
b     L  G  S  Q  L  Q  S  R  S  L  N  G  D  Y  A  K  D  T  A  L  -
c      S  V  A  S  F  S  R  D  L  *  T  A  I  T  R  K  I  P  L  L -
```

```
      TGGTATCGCTGGTAACCAGGCTTCGTCACAGTTGCAGGCCTGGTTACAACATTATGGAAC
781   ----------+---------+---------+---------+---------+---------+ 840
      ACCATAGCGACCATTGGTCCGAAGCAGTGTCAACGTCCGGACCAATGTTGTAATACCTTG
a     W  Y  R  W  *  P  G  F  V  T  V  A  G  L  V  T  T  L  W  N   -
b      G  I  A  G  N  Q  A  S  S  Q  L  Q  A  W  L  Q  H  Y  G  T  -
c       V  S  L  V  T  R  L  R  H  S  C  R  P  G  Y  N  I  M  E  R -


      GGCAGAGGTTAATCTGCAGAGTGGTAATAACTTTGACGGTAGTTCACTGGACTTCTTATT
841   ----------+---------+---------+---------+---------+---------+ 900
      CCGTCTCCAATTAGACGTCTCACCATTATTGAAACTGCCATCAAGTGACCTGAAGAATAA
a     G  R  G  *  S  A  E  W  *  *  L  *  R  *  F  T  G  L  L  I   -
b      A  E  V  N  L  Q  S  G  N  N  F  D  G  S  S  L  D  F  L  L  -
c       Q  R  L  I  C  R  V  V  I  T  L  T  V  V  H  W  T  S  Y  Y -


      ACCGTTCTATGATTCCGAAAAAATGCTGGCATTTGGTCAGGTCGGAGCGCGTTACATTGA
901   ----------+---------+---------+---------+---------+---------+ 960
      TGGCAAGATACTAAGGCTTTTTTACGACCGTAAACCAGTCCAGCCTCGCGCAATGTAACT
a     T  V  L  *  F  R  K  N  A  G  I  W  S  G  R  S  A  L  H  *   -
b      P  F  Y  D  S  E  K  M  L  A  F  G  Q  V  G  A  R  Y  I  D  -
c       R  S  M  I  P  K  K  C  W  H  L  V  R  S  E  R  V  T  L  T -


      CTCCCGCTTTACGGCAAATTTAGGTGCGGGTCAGCGTTTTTTCCTTCCTGCAAACATGTT
961   ----------+---------+---------+---------+---------+---------+ 1020
      GAGGGCGAAATGCCGTTTAAATCCACGCCCAGTCGCAAAAAAGGAAGGACGTTTGTACAA
a     L  P  L  Y  G  K  F  R  C  G  S  A  F  F  P  S  C  K  H  V   -
b      S  R  F  T  A  N  L  G  A  G  Q  R  F  F  L  P  A  N  M  L  -
c       P  A  L  R  Q  I  *  V  R  V  S  V  F  S  F  L  Q  T  C  W -


      GGGCTATAACGTCTTCATTGATCAGGATTTTTCTGGTGATAATACCCGTTTAGGTATTGG
1021  ----------+---------+---------+---------+---------+---------+ 1080
      CCCGATATTGCAGAAGTAACTAGTCCTAAAAAGACCACTATTATGGGCAAATCCATAACC
a     G  L  *  R  L  H  *  S  G  F  F  W  *  *  Y  P  F  R  Y  W   -
b      G  Y  N  V  F  I  D  Q  D  F  S  G  D  N  T  R  L  G  I  G  -
c       A  I  T  S  S  L  I  R  I  F  L  V  I  I  P  V  *  V  L  V -


      TGGCGAATACTGGCGAGACTATTTCAAAAGTAGCGTTAACGGCTATTTCCGCATGAGCGG
1081  ----------+---------+---------+---------+---------+---------+ 1140
      ACCGCTTATGACCGCTCTGATAAAGTTTTCATCGCAATTGCCGATAAAGGCGTACTCGCC
a     W  R  I  L  A  R  L  F  Q  K  *  R  *  R  L  F  P  H  E  R   -
b      G  E  Y  W  R  D  Y  F  K  S  S  V  N  G  Y  F  R  M  S  G  -
c       A  N  T  G  E  T  I  S  K  V  A  L  T  A  I  S  A  *  A  A -
```

```
      CTGGCATGAGTCATACAATAAGAAAGACTATGATGAGCGCCCAGCAAATGGCTTCGATAT
1141  ---------+---------+---------+---------+---------+---------+ 1200
      GACCGTACTCAGTATGTTATTCTTTCTGATACTACTCGCGGGGTCGTTTACCGAAGCTATA
a     L  A  *  V  I  Q  *  E  R  L  *  *  A  P  S  K  W  L  R  Y   -
b      W  H  E  S  Y  N  K  K  D  Y  D  E  R  P  A  N  G  F  D  I  -
c       G  M  S  H  T  I  R  K  T  M  M  S  A  Q  Q  M  A  S  I  S -


      CCGTTTTAATGGCTATCTACCGTCATATCCGGCATTAGGCGCCAAGCTGATATATGAGCA
1201  ---------+---------+---------+---------+---------+---------+ 1260
      GGCAAAATTACCGATAGATGGCAGTATAGGCCGTAATCCGCGGTTCGACTATATACTCGT
a     P  F  *  W  L  S  T  V  I  S  G  I  R  R  Q  A  D  I  *  A   -
b      R  F  N  G  Y  L  P  S  Y  P  A  L  G  A  K  L  I  Y  E  Q  -
c       V  L  M  A  I  Y  R  H  I  R  H  *  A  P  S  *  Y  M  S  S -


      GTATTATGGTGATAATGTTGCTTTGTTTAATTCTGATAAGCTGCAGTCGAATCCTGGTGC
1261  ---------+---------+---------+---------+---------+---------+ 1320
      CATAATACCACTATTACAACGAAACAAATTAAGACTATTCGACGTCAGCTTAGGACCACG
a     V  L  W  *  *  C  C  F  V  *  F  *  *  A  A  V  E  S  W  C   -
b      Y  Y  G  D  N  V  A  L  F  N  S  D  K  L  Q  S  N  P  G  A  -
c       I  M  V  I  M  L  L  C  L  I  L  I  S  C  S  R  I  L  V  R -


      GGCGACCGTTGGTGTAAACTATACTCCGATTCCTCTGGTGACGATGGGGATCGATTACCG
1321  ---------+---------+---------+---------+---------+---------+ 1380
      CCGCTGGCAACCACATTTGATATGAGGCTAAGGAGACCACTGCTACCCCTAGCTAATGGC
a     G  D  R  W  C  K  L  Y  S  D  S  S  G  D  D  G  D  R  L  P   -
b      A  T  V  G  V  N  Y  T  P  I  P  L  V  T  M  G  I  D  Y  R  -
c       R  P  L  V  *  T  I  L  R  F  L  W  *  R  W  G  S  I  T  V -


      TCATGGTACGGGTAATGAAAATGATCTCCTTTACTCAATGCAGTTCCGTTATCAGTTTGA
1381  ---------+---------+---------+---------+---------+---------+ 1440
      AGTACCATGCCCATTACTTTTACTAGAGGAAATGAGTTACGTCAAGGCAATAGTCAAACT
a     S  W  Y  G  *  *  K  *  S  P  L  L  N  A  V  P  L  S  V  *   -
b      H  G  T  G  N  E  N  D  L  L  Y  S  M  Q  F  R  Y  Q  F  D  -
c       M  V  R  V  M  K  M  I  S  F  T  Q  C  S  S  V  I  S  L  I -


      TAAATCGTGGTCTCAGCAAATTGAACCACAGTATGTTAACGAGTTAAGAACATTATCAGG
1441  ---------+---------+---------+---------+---------+---------+ 1500
      ATTTAGCACCAGAGTCGTTTAACTTGGTGTCATACAATTGCTCAATTCTTGTAATAGTCC
a     *  I  V  V  S  A  N  *  T  T  V  C  *  R  V  K  N  I  I  R   -
b      K  S  W  S  Q  Q  I  E  P  Q  Y  V  N  E  L  R  T  L  S  G  -
c       N  R  G  L  S  K  L  N  H  S  M  L  T  S  *  E  H  Y  Q  A -
```


CAGCCGTTACGATCTGGTTCAGCGTAATAACAATATTATTCTGGAGTACAAGAAGCAGGA
1501 ---------+---------+---------+---------+---------+---------+ 1560
GTCGGCAATGCTAGACCAAGTCGCATTATTGTTATAATAAGACCTCATGTTCTTCGTCCT
a Q P L R S G S A * * Q Y Y S G V Q E A G -
b S R Y D L V Q R N N N I I L E Y K K Q D -
c A V T I W F S V I T I L F W S T R S R I -
TATTCTTTCTCTGAATATTCCGCATGATATTAATGGTACTGAACACAGTACGCAGAAGAT
1561 ---------+---------+---------+---------+---------+---------+ 1620
ATAAGAAAGAGACTTATAAGGCGTACTATAATTACCATGACTTGTGTCATGCGTCTTCTA
a Y S F S E Y S A * Y * W Y * T Q Y A E D -
b I L S L N I P H D I N G T E H S T Q K I -
c F F L * I F R M I L M V L N T V R R R F -
TCAGTTGATCGTTAAGAGCAAATACGGTCTGGATCGTATCGTCTGGGATGATAGTGCATT
1621 ---------+---------+---------+---------+---------+---------+ 1680
AGTCAACTAGCAATTCTCGTTTATGCCAGACCTAGCATAGCAGACCCTACTATCACGTAA
a S V D R * E Q I R S G S Y R L G * * C I -
b Q L I V K S K Y G L D R I V W D D S A L -
c S * S L R A N T V W I V S S G M I V H Y -
ACGCAGTCAGGGCGGTCAGATTCAGCATAGCGGAAGCCAAAGCGCACAAGACTACCAGGC
1681 ---------+---------+---------+---------+---------+---------+ 1740
TGCGTCAGTCCCGCCAGTCTAAGTCGTATCGCCTTCGGTTTCGCGTGTTCTGATGGTCCG
a T Q S G R S D S A * R K P K R T R L P G -
b R S Q G G Q I Q H S G S Q S A Q D Y Q A -
c A V R A V R F S I A E A K A H K T T R L -
TATTTTGCCTGCTTATGTGCAAGGTGGCAGCAATATTTATAAAGTGACGGCTCGCGCCTA
1741 ---------+---------+---------+---------+---------+---------+ 1800
ATAAAACGGACGAATACACGTTCCACCGTCGTTATAAATATTTCACTGCCGAGCGCGGAT
a Y F A C L C A R W Q Q Y L * S D G S R L -
b I L P A Y V Q G G S N I Y K V T A R A Y -
c F C L L M C K V A A I F I K * R L A P M -
TGACCGTAATGGCAATAGCTCTAACAATGTACAGCTTACTATTACCGTTCTGTCGAATGG
1801 ---------+---------+---------+---------+---------+---------+ 1860
ACTGGCATTACCGTTATCGAGATTGTTACATGTCGAATGATAATGGCAAGACAGCTTACC
a * P * W Q * L * Q C T A Y Y Y R S V E W -
b D R N G N S S N N V Q L T I T V L S N G -
c T V M A I A L T M Y S L L L P F C R M V -


TCAAGTTGTCGACCAGGTTGGGGTAACGGACTTTACGGCGGATAAGACTTCGGCTAAAGC
1861 1920
AGTTCAACAGCTGGTCCAACCCCATTGCCTGAAATGCCGCCTATTCTGAAGCCGATTTCG
a S S C R P G W G N G L Y G G * D F G * S -
b Q V V D Q V G V T D F T A D K T S A K A -
c K L S T R L G * R T L R R I R L R L K R -
GGATAACGCCGATACCATTACTTATACCGCGACGGTGAAAAAGAATGGGGTAGCTCAGGC
1921 1980
CCTATTGCGGCTATGGTAATGAATATGGCGCTGCCACTTTTTCTTACCCCATCGAGTCCG
a G * R R Y H Y L Y R D G E K E W G S S G -
b D N A D T I T Y T A T V K K N G V A Q A -
c I T P I P L L I P R R * K R M G * L R L -
TAATGTCCCTGTTTCATTTAATATTGTTTCAGGAACTGCAACTCTTGGGGCAAATAGTGC
1981 2040
ATTACAGGGACAAAGTAAATTATAACAAAGTCCTTGACGTTGAGAACCCCGTTTATCACG
a * C P C F I * Y C F R N C N S W G K * C -
b N V P V S F N I V S G T A T L G A N S A -
c M S L F H L I L F Q E L Q L L G Q I V P -
CAAAACGGATGCTAACGGTAAGGCAACCGTAACGTTGAAGTCGAGTACGCCAGGACAGGT
2041 2100
GTTTTGCCTACGATTGCCATTCCGTTGGCATTGCAACTTCAGCTCATGCGGTCCTGTCCA
a Q N G C * R * G N R N V E V E Y A R T G -
b K T D A N G K A T V T L K S S T P G Q V -
c K R M L T V R Q P * R * S R V R Q D R S -
CGTCGTGTCTGCTAAAACCGCGGAGATGACTTCAGCACTTAATGCCAGTGCGGTTATATT
2101 2160
GCAGCACAGACGATTTTGGCGCCTCTACTGAAGTCGTGAATTACGGTCACGCCAATATAA
a R R V C * N R G D D F S T * C Q C G Y I -
b V V S A K T A E M T S A L N A S A V I F -
c S C L L K P R R * L Q H L M P V R L Y F -
TTTTGATCAAACCAAGGCCAGCATTACTGAGATTAAGGCTGATAAGACAACTGCAGTAGC
2161 2220
AAAACTAGTTTGGTTCCGGTCGTAATGACTCTAATTCCGACTATTCTGTTGACGTCATCG
a F * S N Q G Q H Y * D * G * * D N C S S -
b F D Q T K A S I T E I K A D K T T A V A -
c L I K P R P A L L R L R L I R Q L Q * Q -

```
      AAATGGTAAGGATGCTATTAAATATACTGTAAAAGTTATGAAAAACGGTCAGCCAGTTAA
2221  ---------+---------+---------+---------+---------+---------+ 2280
      TTTACCATTCCTACGATAATTTATATGACATTTTCAATACTTTTTGCCAGTCGGTCAATT
a     K  W  *  G  C  Y  *  I  Y  C  K  S  Y  E  K  R  S  A  S  *   -
b      N  G  K  D  A  I  K  Y  T  V  K  V  M  K  N  G  Q  P  V  N  -
c       M  V  R  M  L  L  N  I  L  *  K  L  *  K  T  V  S  Q  L  I -


      TAATCAATCCGTTACATTCTCAACAAACTTTGGGATGTTCAACGGTAAGTCTCAAACGCA
2281  ---------+---------+---------+---------+---------+---------+ 2340
      ATTAGTTAGGCAATGTAAGAGTTGTTTGAAACCCTACAAGTTGCCATTCAGAGTTTGCGT
a     *  S  I  R  Y  I  L  N  K  L  W  D  V  Q  R  *  V  S  N  A   -
b      N  Q  S  V  T  F  S  T  N  F  G  M  F  N  G  K  S  Q  T  Q  -
c       I  N  P  L  H  S  Q  Q  T  L  G  C  S  T  V  S  L  K  R  K -


      AGCAACCACGGGAAATGATGGTCGTGCGACGATAACACTAACTTCCAGTTCCGCCGGTAA
2341  ---------+---------+---------+---------+---------+---------+ 2400
      TCGTTGGTGCCCTTTACTACCAGCACGCTGCTATTGTGATTGAAGGTCAAGGCGGCCATT
a     S  N  H  G  K  *  W  S  C  D  D  N  T  N  F  Q  F  R  R  *   -
b      A  T  T  G  N  D  G  R  A  T  I  T  L  T  S  S  S  A  G  K  -
c       Q  P  R  E  M  M  V  V  R  R  *  H  *  L  P  V  P  P  V  K -


      AGCGACTGTTAGTGCGACAGTCAGTGATGGGGCTGAGGTTAAAGCGACTGAGGTCACTTT
2401  ---------+---------+---------+---------+---------+---------+ 2460
      TCGCTGACAATCACGCTGTCAGTCACTACCCCGACTCCAATTTCGCTGACTCCAGTGAAA
a     S  D  C  *  C  D  S  Q  *  W  G  *  G  *  S  D  *  G  H  F   -
b      A  T  V  S  A  T  V  S  D  G  A  E  V  K  A  T  E  V  T  F  -
c       R  L  L  V  R  Q  S  V  M  G  L  R  L  K  R  L  R  S  L  F -


      TTTTGATGAACTGAAAATTGACAACAAGGTTGATATTATTGGTAACAATGTCAAGAGGTC
2461  ---------+---------+---------+---------+---------+---------+ 2520
      AAAACTACTTGACTTTTAACTGTTGTTCCAACTATAATAACCATTGTTACAGTTCTCCAG
a     F  *  *  T  E  N  *  Q  Q  G  *  Y  Y  W  *  Q  C  Q  E  V   -
b      F  D  E  L  K  I  D  N  K  V  D  I  I  G  N  N  V  K  R  S  -
c       L  M  N  *  K  L  T  T  R  L  I  L  L  V  T  M  S  R  G  R -


      GATGTTGCCTAATATTTGGCTGCAATATGGTCAGTTTAAACTGAAAGCAAGCGGTGGTGA
2521  ---------+---------+---------+---------+---------+---------+ 2580
      CTACAACGGATTATAAACCGACGTTATACCAGTCAAATTTGACTTTCGTTCGCCACCACT
a     D  V  A  *  Y  L  A  A  I  W  S  V  *  T  E  S  K  R  W  *   -
b      M  L  P  N  I  W  L  Q  Y  G  Q  F  K  L  K  A  S  G  G  D  -
c       C  C  L  I  F  G  C  N  M  V  S  L  N  *  K  Q  A  V  V  M -
```

```
     TGGTACATATTCATGGTATTCAGAAAATACCAGTATCGCGACTGTCGATGCATCAGGGAA
2581 ---------+---------+---------+---------+---------+---------+ 2640
     ACCATGTATAAGTACCATAAGTCTTTTATGGTCATAGCGCTGACAGCTACGTAGTCCCTT
a    W  Y  I  F  M  V  F  R  K  Y  Q  Y  R  D  C  R  C  I  R  E   -
b     G  T  Y  S  W  Y  S  E  N  T  S  I  A  T  V  D  A  S  G  K  -
c      V  H  I  H  G  I  Q  K  I  P  V  S  R  L  S  M  H  Q  G  K -


     AGTCACTTTGAATGGTAAAGGCAGTGTCGTAATTAAAGCCACATCTGGTGATAAGCAAAC
2641 ---------+---------+---------+---------+---------+---------+ 2700
     TCAGTGAAACTTACCATTTCCGTCACAGCATTAATTTCGGTGTAGACCACTATTCGTTTG
a    S  H  F  E  W  *  R  Q  C  R  N  *  S  H  I  W  *  *  A  N   -
b     V  T  L  N  G  K  G  S  V  V  I  K  A  T  S  G  D  K  Q  T  -
c      S  L  *  M  V  K  A  V  S  *  L  K  P  H  L  V  I  S  K  Q -


     AGTAAGTTACACTATAAAAGCACCGTCGTATATGATAAAAGTGGATAAGCAAGCCTATTA
2701 ---------+---------+---------+---------+---------+---------+ 2760
     TCATTCAATGTGATATTTTCGTGGCAGCATATACTATTTTCACCTATTCGTTCGGATAAT
a    S  K  L  H  Y  K  S  T  V  V  Y  D  K  S  G  *  A  S  L  L   -
b     V  S  Y  T  I  K  A  P  S  Y  M  I  K  V  D  K  Q  A  Y  Y  -
c      *  V  T  L  *  K  H  R  R  I  *  *  K  W  I  S  K  P  I  M -


     TGCTGATGCTATGTCCATTTGCAAAAATTTATTACCATCCACACAGACGGTATTGTCAGA
2761 ---------+---------+---------+---------+---------+---------+ 2820
     ACGACTACGATACAGGTAAACGTTTTTAAATAATGGTAGGTCTGTCTGCCATAACAGTCT
a    C  *  C  Y  V  H  L  Q  K  F  I  T  I  H  T  D  G  I  V  R   -
b     A  D  A  M  S  I  C  K  N  L  L  P  S  T  Q  T  V  L  S  D  -
c      L  M  L  C  P  F  A  K  I  Y  Y  H  P  H  R  R  Y  C  Q  I -


     TATTTATGACTCATGGGGGGCTGCAAATAAATATAGCCATTATAGTTCTATGAACTCAAT
2821 ---------+---------+---------+---------+---------+---------+ 2880
     ATAAATACTGAGTACCCCCCGACGTTTATTTATATCGGTAATATCAAGATACTTGAGTTA
a    Y  L  *  L  M  G  G  C  K  *  I  *  P  L  *  P  Y  E  L  N   -
b     I  Y  D  S  W  G  A  A  N  K  Y  S  H  Y  S  S  M  N  S  I  -
c      F  M  T  H  G  G  L  Q  I  N  I  A  I  I  V  L  *  T  Q  * -


     AACTGCTTGGATTAAACAGACATCTAGTGAGCAGCGTTCTGGAGTATCAAGCACTTATAA
2881 ---------+---------+---------+---------+---------+---------+ 2940
     TTGACGAACCTAATTTGTCTGTAGATCACTCGTCGCAAGACCTCATAGTTCGTGAATATT
a    N  C  L  D  *  T  D  I  *  *  A  A  F  W  S  I  K  H  L  *   -
b     T  A  W  I  K  Q  T  S  S  E  Q  R  S  G  V  S  S  T  Y  N  -
c      L  L  G  L  N  R  H  L  V  S  S  V  L  E  Y  Q  A  L  I  T -
```

```
     CCTAATAACACAAAACCCTCTTCCTGGGGTTAATGTTAATACTCCAAATGTCTATGCGGT
2941 ---------+---------+---------+---------+---------+---------+ 3000
     GGATTATTGTGTTTTGGGAGAAGGACCCCAATTACAATTATGAGGTTTACAGATACGCCA
a    P  N  N  T  K  P  S  S  W  G  *  C  *  Y  S  K  C  L  C  G  -
b     L  I  T  Q  N  P  L  P  G  V  N  V  N  T  P  N  V  Y  A  V  -
c      *  *  H  K  T  L  F  L  G  L  M  L  I  L  Q  M  S  M  R  F -


     TTGTGTAGAATAATTCCATAACCACCCCGGCTAAAATATGTATTGTTTTAGTCGGGGCAT
3001 ---------+---------+---------+---------+---------+---------+ 3060
     AACACATCTTATTAAGGTATTGGTGGGGCCGATTTTATACATAACAAATCAGCCCCGTA
a    L  C  R  I  I  P  *  P  P  R  L  K  Y  V  L  F  *  S  G  H  -
b     C  V  E  *  F  H  N  H  P  G  *  N  M  Y  C  F  S  R  G  I  -
c      V  *  N  N  S  I  T  T  P  A  K  I  C  I  V  L  V  G  A  * -


     AATTATTTCTTCTTAAGAAATAACCCTCTTATAATCAAATCTACTACTGGTCTTTTTATC
3061 ---------+---------+---------+---------+---------+---------+ 3120
     TTAATAAAGAAGAATTCTTTATTGGGAGAATATTAGTTTAGATGATGACCAGAAAAATAG
a    N  Y  F  F  L  R  N  N  P  L  I  I  K  S  T  T  G  L  F  I  -
b     I  I  S  S  *  E  I  T  L  L  *  S  N  L  L  L  V  F  L  S  -
c      L  F  L  L  K  K  *  P  S  Y  N  Q  I  Y  Y  W  S  F  Y  L -


     TGCTTAATAGG
3121 ---------+- 3131
     ACGAATTATCC

a    C  L  I     -
b     A  *  *     -
c      L  N  R    -
```

8. INFORMATION FOR SEQ ID NO:2:

(a) SEQUENCE CHARACTERISTICS:
(i) LENGTH: 31 base pairs
(ii) TYPE: nucleic acid
(iii) STRANDEDNESS: single
(iv) TOPOLOGY: linear

(b) MOLECULAR TYPE: DNA (genomic)

(c) SEQUENCE DESCRIPTION: SEQ ID NO:2:

5'ATG GAA TTC TCG TCA CAG TTG CAG GCC TGG T 3'

9. INFORMATION FOR SEQ ID NO:3:

(a) SEQUENCE CHARACTERISTICS:
(i) LENGTH: 32 base pairs
(ii) TYPE: nucleic acid
(iii) STRANDEDNESS: single
(iv) TOPOLOGY: linear

(b) MOLECULAR TYPE: DNA (genomic)

(c) SEQUENCE DESCRIPTION: SEQ ID NO:3:

5'ATG GAA TTC CGA AGT CTT ATC AGC CGT AAA GT 3'

10. INFORMATION FOR SEQ ID NO:4:

(a) SEQUENCE CHARACTERISTICS:
(i) LENGTH: 29 base pairs
(ii) TYPE: nucleic acid
(iii) STRANDEDNESS: single
(iv) TOPOLOGY: linear

(b) MOLECULAR TYPE: DNA (genomic)

(c) SEQUENCE DESCRIPTION: SEQ ID NO:4:

5'ATG GAA TTC GGA TGT TCA ACG GTA AGT CT 3'

11. INFORMATION FOR SEQ ID NO:5:

(a) SEQUENCE CHARACTERISTICS:
(i) LENGTH: 29 base pairs
(ii) TYPE: nucleic acid
(iii) STRANDEDNESS: single
(iv) TOPOLOGY: linear

(b) MOLECULAR TYPE: DNA (genomic)

(c) SEQUENCE DESCRIPTION: SEQ ID NO:5:

5'ATG GAA TTC ACA ATA CCG TCT GTG TGG AT 3'

12. INFORMATION FOR SEQ ID NO:6:

(a) SEQUENCE CHARACTERISTICS:
(i) LENGTH: 17 base pairs
(ii) TYPE: nucleic acid
(iii) STRANDEDNESS: single
(iv) TOPOLOGY: linear

(b) MOLECULAR TYPE: DNA (genomic)

(c) SEQUENCE DESCRIPTION: SEQ ID NO:6:

5'AAG CGA CTG AGG TCA CT 3'

13. INFORMATION FOR SEQ ID NO:7:

(a) SEQUENCE CHARACTERISTICS:
(i) LENGTH: 18 base pairs
(ii) TYPE: nucleic acid
(iii) STRANDEDNESS: single
(iv) TOPOLOGY: linear

(b) MOLECULAR TYPE: DNA (genomic)

(c) SEQUENCE DESCRIPTION: SEQ ID NO:7:

5'ACG CTG CTC ACT AGA TGT 3'

14. INFORMATION FOR SEQ ID NO:8:

(a) SEQUENCE CHARACTERISTICS:
(i) LENGTH: 17 base pairs
(ii) TYPE: nucleic acid
(iii) STRANDEDNESS: single
(iv) TOPOLOGY: linear

(b) MOLECULAR TYPE: DNA (genomic)

(c) SEQUENCE DESCRIPTION: SEQ ID NO:8:

5'AGA CCT AGG TTA CAT TT 3'

15. INFORMATION FOR SEQ ID NO:9:

(a) SEQUENCE CHARACTERISTICS:
(i) LENGTH: 18 base pairs
(ii) TYPE: nucleic acid
(iii) STRANDEDNESS: single
(iv) TOPOLOGY: linear

(b) MOLECULAR TYPE: DNA (genomic)

(c) SEQUENCE DESCRIPTION: SEQ ID NO:8:

5'TAT TTT ATC AGC TTC AGT 3'

THE EMBODIMENTS OF THE INVENTION IN WHICH AN EXCLUSIVE PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:

1. A purified and isolated DNA segment having a sequence which codes for a protein associated with attaching and effacing activity of enterohemorrhagic E.coli, or an oligonucleotide fragment of the DNA segment.

2. A purified and isolated DNA segment as claimed in claim 1 having a sequence which codes for a protein associated with attaching and effacing activity of enterohemorrhagic E.coli which has substantial homology with the amino acid sequence as shown in the Sequence Listing as SEQ ID NO:1.

3. A purified and isolated DNA segment as claimed in claim 1 having a sequence having substantial sequence homology with the nucleotide sequence as shown in the Sequence Listing as SEQ ID NO:1.

4. A purified and isolated DNA segment as claimed in claim 1 having a sequence having more than ? sequence homology with the nucleotide sequence as shown in the Sequence Listing as SEQ ID NO:1.

5. A purified and isolated single stranded nucleotide sequence comprising a DNA segment as claimed in claim 2.

6. An antisense sequence of a DNA segment as claimed in claim 1 or an oligonucleotide fragment of the antisense sequence.

7. A recombinant molecule adapted for transformation of a host cell comprising a DNA segment or an oligonucleotide fragment of the DNA segment as claimed in claim 1 and an expression control sequence operatively linked to the DNA segment.

8. A transformant host cell including a recombinant molecule as claimed in claim 7.

9. A method for preparing a protein associated with attaching and effacing activity of enterohemorrhagic E.coli utilizing a purified and isolated DNA segment as claimed in claim 2.

10. A substantially pure protein associated with attaching and effacing activity of enterohemorrhagic E.coli.

11. A substantially pure protein as claimed in claim 10 having an amino acid sequence which has substantial homology with the amino acid sequence as shown in the Sequence Listing as SEQ ID NO:1.

12. A protein encoded by the purified and isolated DNA segment as claimed in claim 2.

13. A monoclonal or polyclonal antibody specific for an epitope of a protein as claimed in claim 10.

14. An antibody as claimed in claim 13 which binds a distinct epitope in an unconserved region of the protein.

15. A purified and isolated nucleotide sequence comprising the sequence shown in the Sequence Listing as SEQ ID NO:2 or the sequence shown in the Sequence Listing as SEQ ID NO:3, or a sequence having substantial homology thereto or a fragment of the nucleotide sequence.

16. A purified and isolated nucleotide sequence comprising the sequence shown in the Sequence Listing as SEQ ID NO:4 or the sequence shown in the Sequence Listing as SEQ ID NO:5, or a sequence having substantial homology thereto or a fragment of the nucleotide sequence.

17. A purified and isolated nucleotide sequence comprising the sequence shown in the Sequence Listing as SEQ ID NO:6 or the sequence shown in the Sequence Listing as SEQ ID NO:7, or a sequence having substantial homology thereto or a fragment of the nucleotide sequence.

18. A purified and isolated nucleotide sequence comprising the sequence shown in the Sequence Listing as SEQ ID NO:8 or the sequence shown in the Sequence Listing as SEQ ID NO:9, or a sequence having substantial homology thereto or a fragment of the nucleotide sequence.

19. A diagnostic kit for detecting enterohemorrhagic E.coli 0157 in a sample comprising a monoclonal antibody as claimed in claim 13 and directions for its use.

20. An immunoassay for the detection of E.coli 0157 comprising contacting a sample suspected of containing E.coli 0157 with a monoclonal antibody as claimed in claim 13 to form an immune complex; and determining the presence of the immune complex in order to detect E.coli 0157 in the sample.

21. An avirulent strain of an enterohemorrhagic E.coli comprising an avirulent bacterial carrier strain transformed with a recombinant molecule as claimed in claim 7.

22. A vaccine composition comprising a bacterial carrier strain transformed with a recombinant molecule as claimed in claim 7.

23. A method for detecting the presence of a DNA segment having a sequence encoding a protein related to or analogous to a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or an oligonucleotide fragment thereof in a sample, comprising contacting the sample with a nucleotide probe capable of hybridizing with the DNA segment or an oligonucleotide fragment thereof, to form a hybridization product, under conditions which permit the formation of the hybridization product, and assaying for the hybridization product.

24. A kit for detecting the presence of a DNA segment having a sequence encoding a protein related to or analogous to a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or an oligonucleotide fragment thereof,

in a sample comprising a nucleotide probe capable of hybridizing with the DNA segment or an oligonucleotide fragment thereof, reagents required for hybridization of the nucleotide probe with the DNA segment or an oligonucleotide fragment thereof, and directions for its use.

25. A method of determining the presence of a DNA segment having a sequence encoding a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or a predetermined oligonucleotide fragment thereof in a sample, comprising treating the sample with primers which are capable of amplifying the DNA segment or the predetermined oligonucleotide fragment thereof in a polymerase chain reaction to form amplified sequences, under conditions which permit the formation of amplified sequences, and assaying for amplified sequences.

26. A kit for determining the presence of a DNA segment having a sequence encoding a protein associated with attaching and effacing activity of enterohemorrhagic E.coli or a predetermined oligonucleotide fragment thereof in a sample, comprising primers which are capable of amplifying the DNA segment or the predetermined oligonucleotide fragment thereof in a polymerase chain reaction to form amplified sequences, means for assaying the amplified sequences, and directions for its use.

27. A method for detecting verotoxin-producing E.coli of the serogroups 05, 026, 0103, 0111, 0118, 0145, and 0157 in a sample comprising contacting the sample with a SalI-StuI fragment (C1-C2) comprising the sequence as shown in the Sequence Listing as SEQ ID NO:2 and the sequence as shown in the Sequence Listing as SEQ ID NO:3, under conditions which permit the fragment to hybridize with a complementary sequence in the sample to form a hybridization product, and assaying for the hybridization product.

28. A method for detecting verotoxin-producing E.coli serogroup 0157 in a sample comprising contacting the sample with a 0.5 kb fragment (A3-B2) comprising the sequence as shown in the Sequence

Listing as SEQ ID NO:4 and the sequence as shown in the Sequence Listing as SEQ ID NO:5, under conditions which permit the fragment to hybridize with a complementary sequence in the sample to form a hybridization product, and assaying for the hybridization product.

29. A method for detecting pathogenic verotoxin-producing E.coli strains of the serogroup 0157 in a sample comprising treating the sample with the nucleotide sequence P1EH shown in the Sequence Listing as SEQ ID NO:4 and the nucleotide sequence P2EH5 shown in the Sequence Listing as SEQ ID NO:5, in the polymerase chain reaction, under conditions which permit the formation of amplified sequences which terminate with the nucleic acid sequence of one of the primers and the complementary nucleic acid sequence of the other of the primers; and, assaying for the amplified sequences.

30. A kit for detecting verotoxin-producing E.coli of the serogroups 05, 026, 0103, 0111, 0118, 0145, and 0157 in a sample comprising a SalI-StuI fragment (C1-C2) comprising a sequence as shown in the Sequence Listing as SEQ ID NO:2 and the sequence as shown the Sequence Listing as SEQ ID NO:3, reagents required for hybridization of the fragment with a DNA segment having a sequence encoding the attaching and effacing protein of enterohemorrhagic E.coli or an oligonucleotide fragment thereof, and directions for its use.

31. A kit for detecting verotoxin-producing E.coli of the serogroup 0157 in a sample comprising a 0.5 kb fragment (A3-B2) comprising a sequence as shown in the Sequence Listing as SEQ ID NO:3 and the sequence as shown in the Sequence Listing as SEQ ID NO:4, reagents required for hybridization of the fragment with a DNA segment encoding the attaching and effacing protein of enterohemorrhagic E.coli or an oligonucleotide fragment thereof, and directions for its use.

32. A kit for detecting pathogenic verotoxin-producing E.coli strains of the serogroup 0157 in a sample comprising the primers

P1EH as shown in the Sequence Listing as SEQ ID NO:5 and P2EH as shown in the Sequence Listing as SEQ ID NO:6, reagents required for the amplification in a PCR reaction of sequences which terminate with the nucleic acid sequence of one of the primers and the complementary nucleic acid sequence of the other of the primers, means for assaying amplified sequences, and directions for its use.

2078716

FIGURE 1

```
1     TCGAGAATGAAATAGAAGTCGTTGTTAAGTCAATGGAAAACCTGTATTTGGTATTACATAATCAGGGAATAACATTAGAAAACGAACATATGAAAATAGA  100
101   GGAAATCAGTTCAAGCGACAATAAACATTATTACGCCGGAAGATAAAATCCGATCTATTAATATAATTTATTTCTCATTCTAACTCATTGTGGTGGAGCC  200
        M  I  T  H  G  C  Y  T  R  T  R  H  K  H  K  L  K  K  T  L  I  M  L  S  A  G  L  G  L  F  F  Y
201   ATAACATGATTACTCATGGTTGTTATACCCGGACCCGGCACAAGCATAAGCTAAAAAAAACATTGATTATGCTTAGTGCTGGTTTAGGATTGTTTTTTTA  300
       V  N  Q  N  S  F  A  N  G  E  N  Y  F  K  L  G  S  D  S  K  L  L  T  H  D  S  Y  Q  N  R  L  F  Y
301   TGTTAATCAGAATTCATTTGCAAATGGTGAAAATTATTTTAAATTGGGTTCGGATTCAAAACTGTTAACTCATGATAGCTATCAGAATCGCCTTTTTTAT  400
      T  L  K  T  G  E  T  V  A  D  L  S  K  S  Q  D  I  N  L  S  T  I  W  S  L  N  K  H  L  Y  S  S  E  S
401   ACGTTGAAAACTGGTGAAACTGTTGCCGATCTTTCTAAATCGCAAGATATTAATTTATCGACGATTTGGTCGTTGAATAAGCATTTATACAGTTCTGAAA  500
        E  M  M  K  A  A  P  G  Q  Q  I  I  L  P  L  K  K  L  P  F  E  Y  S  A  L  P  L  L  G  S  A  P  L
501   GCGAAATGATGAAGGCCGCGCCTGGTCAGCAGATCATTTTGCCACTCAAAAAACTTCCCTTTGAATACAGTGCACTACCACTTTTAGGTTCGGCACCTCT  600
       V  A  A  G  G  V  A  G  H  T  N  K  L  T  K  M  S  P  D  V  T  K  S  N  M  T  D  D  K  A  L  N  Y
601   TGTTGCTGCAGGTGGTGTTGCTGGTCACACGAATAAACTGACTAAAATGTCCCCGGACGTGACCAAAAGCAACATGACCGATGACAAGGCATTAAATTAT  700
      A  A  Q  Q  A  A  S  L  G  S  Q  L  Q  S  R  S  L  N  G  D  Y  A  K  D  T  A  L  G  I  A  G  N  Q  A
701   GCGGCACAACAGGCGGCGAGTCTCGGTAGCCAGCTTCAGTCGCGATCTCTGAACGGCGATTACGCGAAAGATACCGCTCTTGGTATCGCTGGTAACCAGG  800
        S  S  Q  L  Q  A  W  L  Q  H  Y  G  T  A  E  V  N  L  Q  S  G  N  N  F  D  G  S  S  L  D  F  L  L
801   CTTCGTCACAGTTGCAGGCCTGGTTACAACATTATGGAACGGCAGAGGTTAATCTGCAGAGTGGTAATAACTTTGACGGTAGTTCACTGGACTTCTTATT  900
       P  F  Y  D  S  E  K  N  L  A  F  G  Q  V  G  A  R  Y  I  D  S  R  F  T  A  N  L  G  A  G  Q  R  F
901   ACCGTTCTATGATTCCGAAAAAAATCTGGCATTTGGTCAGGTCGGAGCGCGTTACATTGACTCCCGCTTTACGGCAAATTTAGGTGCGGGTCAGCGTTTT  1000
      F  L  P  A  N  M  L  G  Y  N  V  F  I  D  Q  D  F  S  G  D  N  T  R  L  G  I  G  G  E  T  W  R  D  Y
1001  TTCCTTCCTGCAAACATGTTGGGCTATAACGTCTTCATTGATCAGGATTTTTCTGGTGATAATACCCGTTTAGGTATTGGTGGCGAAACTTGGCGAGACT  1100
       F  K  S  S  V  N  G  Y  F  R  M  S  G  W  H  E  S  Y  N  K  K  D  Y  D  E  R  P  A  N  G  F  D  I
1101  ATTTCAAAAGTAGCGTTAACGGCTATTTCCGCATGAGCGGCTGGCATGAGTCATACAATAAGAAAGACTATGATGAGCGCCCAGCAAATGGCTTCGATAT  1200
       R  F  N  G  Y  L  P  S  Y  P  A  L  G  A  K  L  I  Y  E  Q  Y  Y  G  D  N  V  A  L  F  N  S  D  K
1201  CCGTTTTAATGGCTATCTACCGTCATATCCGGCATTAGGCGCCAAGCTGATATATGAGCAGTATTATGGTGATAATGTTGCTTTGTTTAATTCTGATAAG  1300
      L  Q  S  N  P  G  A  A  T  V  G  V  N  Y  T  P  I  P  L  V  T  M  G  I  D  Y  R  H  G  T  G  N  E  N
1301  CTGCAGTCGAATCCTGGTGCGGCGACCGTTGGTGTAAACTATACTCCGATTCCTCTGGTGACGATGGGGATCGATTACCGTCATGGTACGGGTAATGAAA  1400
        D  L  L  Y  S  M  Q  F  R  Y  Q  F  D  K  S  W  S  Q  Q  I  E  P  Q  Y  V  N  E  L  R  T  L  S  G
1401  ATGATCTCCTTTACTCAATGCAGTTCCGTTATCAGTTTGATAAATCGTGGTCTCAGCAAATTGAACCACAGTATGTTAACGAGTTAAGAACATTATCAGG  1500
       S  R  Y  D  L  V  Q  R  N  N  N  I  I  L  E  Y  K  K  Q  D  I  L  S  L  N  I  P  H  D  I  N  G  T
1501  CAGCCGTTACGATCTGGTTCAGCGTAATAACAATATTATTCTGGAGTACAAGAAGCAGGATATTCTTTCTCTGAATATTCCGCATGATATTAATGGTACT  1600
      E  H  S  T  Q  K  I  G  L  I  V  K  S  K  Y  G  L  D  R  I  V  W  D  D  S  A  L  R  S  Q  G  G  Q  I
1601  GAACACAGTACGCAGAAGATTGGTCTGATCGTTAAGAGCAAATACGGTCTGGATCGTATCGTCTGGGATGATAGTGCATTACGCAGTCAGGGCGGTCAGA  1700
        Q  H  S  G  S  Q  S  A  Q  D  Y  Q  A  I  L  P  A  Y  V  Q  G  G  S  N  I  Y  K  V  T  A  R  A  Y
1701  TTCAGCATAGCGGAAGCCAAAGCGCACAAGACTACCAGGCTATTTTGCCTGCTTATGTGCAAGGTGGCAGCAATATTTATAAAGTGACGGCTCGCGCCTA  1800
       D  R  N  G  N  S  S  N  N  V  Q  L  T  I  T  V  L  S  N  G  Q  V  V  D  Q  V  G  V  T  D  F  T  A
1801  TGACCGTAATGGCAATAGCTCTAACAATGTACAGCTTACTATTACCGTTCTGTCGAATGGTCAAGTTGTCGACCAGGTTGGGGTAACGGACTTTACGGCG  1900
      D  K  T  S  A  K  A  D  N  A  D  T  I  T  Y  T  A  T  V  K  K  N  G  V  A  Q  A  N  V  P  V  S  F  N
1901  GATAAGACTTCGGCTAAAGCGGATAACGCCGATACCATTACTTATACCGCGACGGTGAAAAAGAATGGGGTAGCTCAGGCTAATGTCCCTGTTTCATTTA  2000
        I  V  S  G  T  A  T  L  G  A  N  S  A  K  T  D  A  N  G  K  A  T  V  T  L  K  S  S  T  P  G  Q  V
2001  ATATTGTTTCAGGAACTGCAACTCTTGGGGCAAATAGTGCCAAAACGGATGCTAACGGTAAGGCAACCGTAACGTTGAAGTCGAGTACGCCAGGACAGGT  2100
       V  V  S  A  K  T  A  E  M  T  S  A  L  N  A  S  A  V  I  F  F  D  Q  T  K  A  S  I  T  E  I  K  A
2101  CGTCGTGTCTGCTAAAACCGCGGAGATGACTTCAGCACTTAATGCCAGTGCGGTTATATTTTTTGATCAAACCAAGGCCAGCATTACTGAGATTAAGGCT  2200
      D  K  T  T  A  V  A  N  G  K  D  A  I  K  T  T  V  K  V  M  K  N  G  Q  P  V  N  N  Q  S  V  T  F  S
2201  GATAAGACAACTGCAGTAGCAAATGGTAAGGATGCTATTAAAACTACTGTAAAAGTTATGAAAAACGGTCAGCCAGTTAATAATCAATCCGTTACATTCT  2300
        T  N  F  G  M  F  N  G  K  S  Q  T  Q  A  T  T  G  N  D  G  R  A  T  I  T  L  T  S  S  S  A  G  K
2301  CAACAAACTTTGGGATGTTCAACGGTAAGTCTCAAACGCAAGCAACCACGGGAAATGATGGTCGTGCGACGATAACACTAACTTCCAGTTCCGCCGGTAA  2400
       A  T  V  S  A  T  V  S  D  G  A  E  V  K  A  T  E  V  T  F  F  D  E  L  K  I  D  N  K  V  D  I  I
2401  AGCGACTGTTAGTGCGACAGTCAGTGATGGGGCTGAGGTTAAAGCGACTGAGGTCACTTTTTTTGATGAACTGAAAATTGACAACAAGGTTGATATTATT  2500
      G  N  N  V  K  R  S  M  L  P  N  I  W  L  Q  Y  G  Q  F  K  L  K  A  S  G  G  D  G  T  Y  S  W  Y  S
2501  GGTAACAATGTCAAGAGGTCGATGTTGCCTAATATTTGGCTGCAATATGGTCAGTTTAAACTGAAAGCAAGCGGTGGTGATGGTACATATTCATGGTATT  2600
       E  N  T  S  I  A  T  V  D  A  S  G  K  V  T  L  N  G  K  G  S  V  V  I  K  A  T  S  G  D  K  Q  T
2601  CAGAAAATACCAGTATCGCGACTGTCGATGCATCAGGGAAAGTCACTTTGAATGGTAAAGGCAGTGTCGTAATTAAAGCCACATCTGGTGATAAGCAAAC  2700
       V  S  Y  T  I  K  A  P  S  Y  M  I  K  V  D  K  Q  A  Y  Y  A  D  A  M  S  I  C  K  N  L  L  P  S
2701  AGTAAGTTACACTATAAAAGCACCGTCGTATATGATAAAAGTGGATAAGCAAGCCTATTATGCTGATGCTATGTCCATTTGCAAAAATTTATTACCATCC  2800
      T  Q  T  V  L  S  D  I  Y  D  S  W  G  A  A  N  K  Y  S  H  Y  S  S  M  N  S  I  T  A  W  I  K  Q  T
2801  ACACAGACGGTATTGTCAGATATTTATGACTCATGGGGGGCTGCAAATAAATATAGCCATTATAGTTCTATGAACTCAATAACTGCTTGGATTAAACAGA  2900
        S  S  E  Q  R  S  G  V  S  S  T  Y  N  L  I  T  Q  N  P  L  P  G  V  N  V  N  T  P  N  V  Y  A  V
2901  CATCTAGTGAGCAGCGTTCTGGAGTATCAAGCACTTATAACCTAATAACACAAAACCCTCTTCCTGGGGTTAATGTTAATACTCCAAATGTCTATGCGGT  3000
        C  V  E  *
3001  TTGTGTAGAATAATTCCATAACCACCCCGGCTAAAATATGTATTGTTTTAGTCGGGGCATAATTATTTCTTCTTAAGAAATAACCCTCTTATAATCAAAT  3100
3101  CTACTACTGGTCTTTTTATCTGCTTAATAGG  3131
```

FIGURE 2

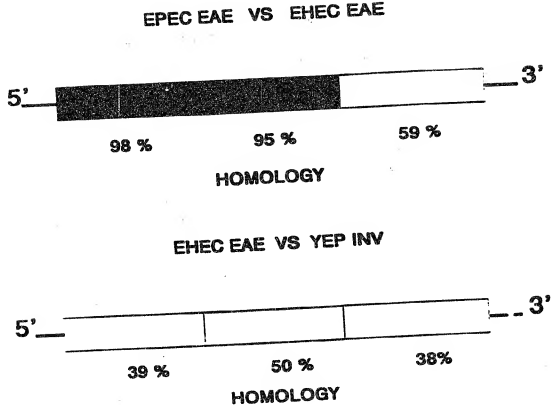

EPEC EAE VS EHEC EAE
5'
3'
98 %
95 %
59 %
HOMOLOGY
EHEC EAE VS YEP INV
5'
3'
39 %
50 %
38%
HOMOLOGY

FIGURE 3

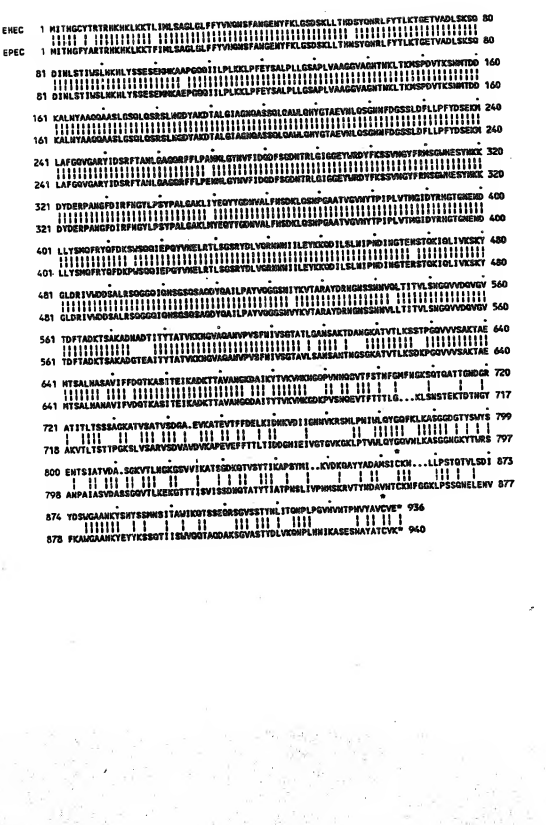


FIGURE 4

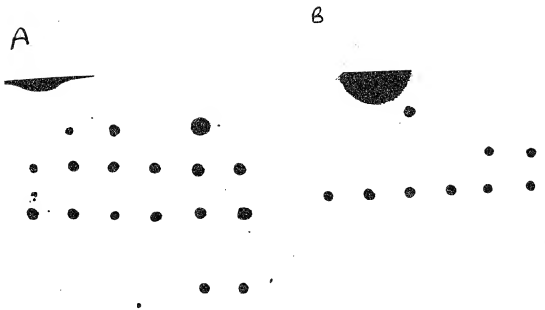


FIGURE 5

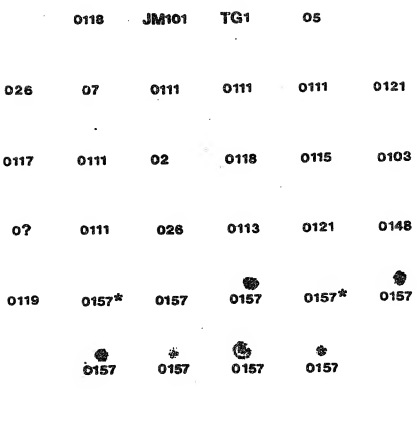

0118
JM101
TG1
05
026
07
0111
0111
0111
0121
0117
0111
02
0118
0115
0103
0?
0111
026
0113
0121
0148
0119
0157*
0157
0157
0157*
0157
0157
0157
0157
0157

FIGURE 6

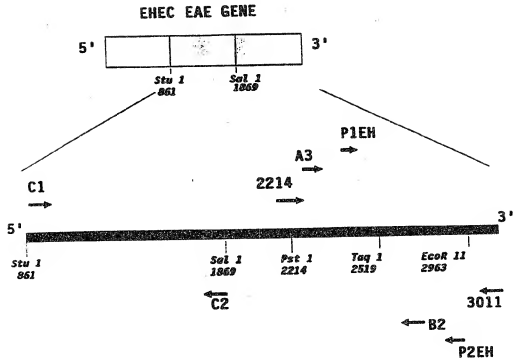

EHEC EAE GENE
5'
3'
Stu 1
861
Sal 1
1869
P1EH
A3
2214
C1
5'
3'
Stu 1
861
Sal 1
1869
Pst 1
2214
Taq 1
2519
EcoR 11
2963
C2
3011
B2
P2EH

FIGURE 7

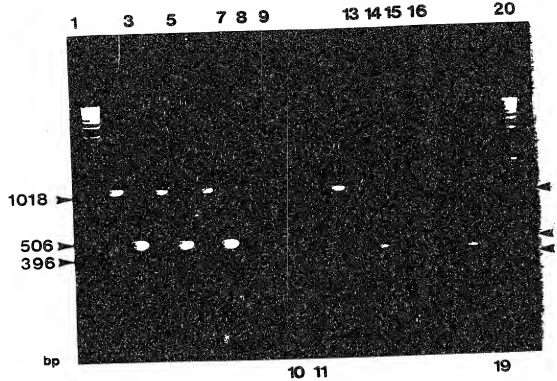


FIGURE 8

FIGURE 9

O157: EHEC eae O157:H7 protein (3)
O111: EHEC eae O111:H8 protein
EPEC: EPEC eae O127:H6 protein (1)
INV : Y.pseudotuberculosis invasin (5)

```
O157  824  IKATSGDKQTVSYTIKAPSYMI..KVDKQAYYADANSICKN...LLPST  867
O111  824  SNKLHCDKQTVSYTIKLRYYMI..RVGNKASYANAMSFCGN...LLPSA  867
EPEC  822  ISVISSDNQTATYTIATPNSLIVPNMSKRVTYNDAVNTCKNFGGKLPSS  870
INV   865  VAVTAKSKKFPSYSV...SYRFYPNRW...IY.DGGRSLVSSLEASRQC  906


O157  868  QTVLSD.IYDSWGAANK.........YS......HYSS.MNSITAWIKQ  899
O111  868  QTVLSN.VYNSWGPANG.........YD......HYRS.MQSITAWITQ  899
EPEC  871  QNELSN.VFKAWGAANK.........YE......YYKS.SQTIISWVQQ  902
INV   907  QGSDMSAVLESSRATNGTRAPDGTLWGEWGSLT.AYSSDWQSGEYWVKK  954


O157  900  TSSEQRSGVSSTYNLITQNPLPGVNV  925
O111  900  TEADKISGVSTTYDLITQNPHKDVLH  925
EPEC  903  TAQDAKSGVASTYDLVKQNPLNNIKA  928
INV   955  TSTDFETNNMDTGAL..Q.PGPAYLA  977
```

2078716